THORSTEN SCHMIDT

WAHRSCHEINLICHKEITS-THEORIE

UNIVERSITÄT FREIBURG

Einführung

Diese Vorlesung ist eine Fortführung der Vorlesung Wahrscheinlichkeitstheorie aus dem Wintersemester 17/18 und aus dem Sommersemester 2022. Nun schließt die Vorlesung direkt an die Stochastik I an, so dass das Skriptum deutlich verändert wird. Das alte Skriptum kann aber als gute Ergänzung zu diesem genutzt werden.

Dieses Skriptum ist vorläufig ! Es enthält offensichtlich noch viele Fehler – also bitte mit dem entsprechenden Fingerspitzengefühl lesen. Für Rückmeldungen sind wir natürlich äußerst dankbar und würden uns über eine Email an das Sekretariat Stochastik.

PS: Auf dem Titelblatt eine zwei-dimensionale Brownsche Bewegung, erstellt mit dem im wesentlichen drei Zeilen umfassenden R-Code aus Beispiel 83.

Thorsten Schmidt

Inhaltsverzeichnis

Einführung 3

Maßtheorie 9

Motivation und Überblick 10

Elementare Strukturen 10

Maße 11

Maße auf $\mathbb{R}$ 12

Approximation von messbaren Funktionen 14

Konvergenzsätze für das Integral 15

Die L^p-Räume 17

Wahrscheinlichkeitstheorie 19

Grundlagen 19

Der Erwartungswert. 20

Momente. 20

Beispiele 20

Die von X erzeugte σ-Algebra 21

Wichtige Ungleichungen 22

Charakteristische Funktionen 24

Konvergenzen 27

Fast sichere und L^p-Konvergenz 31

Das starke Gesetz der großen Zahl 37
Unabhängigkeit und 0-1 Gesetze 37
Terminale Ereignisse 39
Die Maximalungleichung 42
Der Drei-Reihensatz 44
Das starke Gesetz der großen Zahl 45
Das Glivenko-Cantelli Lemma 48
Signierte Maße und der Satz von Radon-Nikodým 50
Produkträume 54
Der Satz von Fubini 57
Die Faltung 60
Der Existenzsatz von Kolmogorov 61

Schwache Konvergenz 65
Schwache Konvergenz 65
Schwache Konvergenz auf $\mathbb{R}$ 71
Straffheit und relative Kompaktheit 74
Kompaktheit und verwandte Begriffe 76
Der Satz von Prohorov 79

Der Zentrale Grenzwertsatz 83
Charakteristische Funktionen 83
Der Satz von Lévy 85
Der zentrale Grenzwertsatz von Lindeberg und Feller 89
Konvergenz des Binomialmodells gegen das Black-Scholes Modell 95
Handelsstrategien 96
Mehrdimensionale Grenzwertsätze 98

Martingale 101
Bedingte Erwartung 101

Stochastische Prozesse und Martingale 108
Wichtige Ungleichungen 113
Doob's Martingalkonvergenzsatz 117
Rückwärtsmartingale 121
Das Fundamental Theorem of Asset Pricing 124

Anhang: Maßtheorie 125
Elementare Strukturen 125
Maße 128
Die Fortsetzung von Maßen 133
Maße auf $\mathbb{R}$ 137
Hilberträume 142
Signierte Maße und der Satz von Radon-Nikodým 143
Produkträume 149
Der Satz von Fubini 152
Der Existenzsatz von Kolmogorov 156

Sonstiges 161
Das Banach-Tarski Paradoxon 161

Literaturverzeichnis 163

Ein sehr kurzer Ausflug in die Maßtheorie

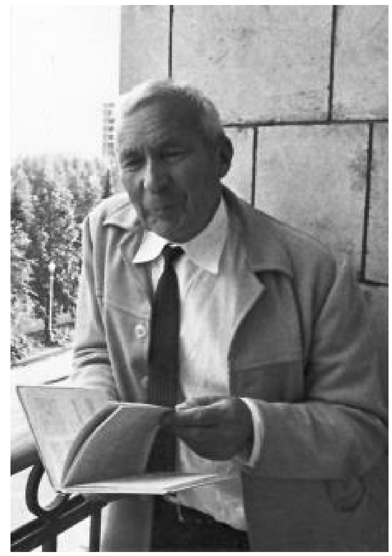


Abbildung 1: Andrej Nikoljewitsch Kolmogorov - Foto by Konrad Jacobs

Im Gegensatz aus der Vorlesung vom letzten Jahr soll dieses Jahr die Maßtheorie nur einen kleinen Teil der Vorlesung einnehmen. Die detaillierte Ausarbeitung wurde deswegen in den Anhang verschoben und das Skriptum startet nur mit einem kurzen Ausschnitt.

Zentrale Aussagen dieses Kapitels:

- Grundlegende Definitionen.
- Darstellung eines Wahrscheinlichkeitsmaßes durch seine Verteilungsfunktion
- Hilfsmittel zur Messbarkeit, Erzeugendensysteme, Messbarkeit von Kompositionen von Abbildungen
- Approximation von messbaren Abbildungen
- Das Integral und zugehörige Regeln
- Die Konvergenzsätze
- L^p-Räume und Konvergenz in L^p.

Die Entwicklung eines präzisen Begriffs für Wahrscheinlichkeit hat die Mathematiker sehr lange beschäftigt. Es war die Idee von Andrei N. Kolmogorov (1903–1987) das Hilfsmittel „*Maß*" aus der Analysis hierfür zu verwenden und Zugang zu der mächtigen Maß-Integrationstheorie von Henri L. Lebesgue (1875–1941) zu erlangen.

Aus diesem Grund sind die Begriffe *Maß* und *Integral* für Stochastiker besonders wichtig. Im Unterschied zur Analysis, beziehungsweise zum Lebesgue-Maß auf $\mathbb{R}^n$, sind wir allerdings stets an endlichen Maßen ($\mu(\Omega) = 1$) interessiert, aber mit einem beliebigen Ω.

Motivation und Überblick

Das klassische Beispiel einer σ-Algebra ist die Borel σ-Algebra $\mathscr{B}(\mathbb{R})$. Sie wird von abzählbar vielen Mengen erzeugt: etwa von den Intervallen $(a,b]$ mit rationalen Intervallgrenzen. Diese Menge ist kein Ring, sondern lediglich ein Halbring[1]. Wir werden hierauf im Kerntext nicht genau eingehen, aber im Anhang die komplette Theorie entwickeln.

[1] Siehe Anhang, Definition A.8.

Für die Erzeugung der Borel σ-Algebra muss man eigentlich lediglich wissen, was offene Mengen sind, denn $\mathscr{B}(\mathbb{R})$ wird auch von allen offenen Mengen erzeugt. Man kan demnach allgemein auf topologischer Räumen arbeiten. Ein *topologischer Raum*[2] $(\Omega, \mathcal{O})$ ist ein Raum Ω versehen mit einer Menge $\mathcal{O} \subset \mathscr{P}(\Omega)$, wobei die Elemente von $\mathcal{O}$ als *offen* bezeichnet werden, so dass

[2] $\leftarrow$ topologischer Raum $(\Omega, \mathcal{O})$

(i) $\varnothing, \Omega \in \mathcal{O}$
(ii) der endliche Schnitt von offenen Mengen ist wieder offen
(iii) die beliebige Vereinigung von offenen Mengen ist wieder offen.

Hat man z.B. einen metrischen Raum (etwa die stetigen Funktionen über $[0,1]$ versehen mit der Supremumsnorem), so induziert dies eine *Topologie* über die Menge der offenen Kugeln.

Elementare Strukturen

Sei Ω eine (beliebige) Menge. Betrachten wir für einen Moment $\Omega = [0,1]$. Wir ordnen jedem Intervall die Wahrscheinlichkeit $P((a,b]) = b - a$ zu. Angenommen [3] wir wollten dieses Maß auf alle Teilmengen von $[0,1]$ ausdehnen, so dass (i) $P(\Omega) = 1$, P ist σ-additiv für paarweise disjunkte Mengen, so kann man zeigen, dass keine solche Erweiterung existiert! Es sind einfach zu viele Mengen. Émile Borel entdeckte, dass wir dies aber auf einer kleineren Menge tun können. Mit $\mathscr{P}(\Omega)$ bezeichnen wir die Menge aller Teilmengen von Omega.

[3] Dies führt zum berühmten Banach-Tarski Paradoxon.

Definition 1. Eine Menge $\mathscr{F} \subset \mathscr{P}(\Omega)$ heißt *σ-Algebra*, falls für alle $A, A_1, A_2, \ldots \in \mathscr{F}$ gilt, dass

(i) $\varnothing \in \mathscr{F}$
(ii) $A^c \in \mathscr{F}$
(iii) $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$

Es ist sehr nützlich die von einer bestimmten Menge erzeugte σ-Algebra zu betrachten (etwa die von den Halbstrahlen erzeugte Borel σ-Algebra). Die grundlegende Eigenschaft hierfür ist, dass der

Schnitt von (beliebig vielen) σ-Algebren wieder eine σ-Algebra ($\rightarrow$ Übung) ist.

Dann[4] kann man für ein $C \subseteq \mathscr{P}(\Omega)$ die von C **erzeugte σ-Algebra** definieren durch

$$\sigma(C) = \bigcap \{\mathscr{F} \supseteq C : \mathscr{F} \text{ ist } \sigma\text{-Algebra}\}. \tag{2}$$

[4] Ein wichtiges Konzept: Wir erzeugen eine σ-Algebra aus einer deutlich einfacheren Menge C.

Definition 3. Sei $(\Omega, \mathcal{O})$ ein topologischer Raum. Dann heißt $\mathscr{B}(\Omega) = \sigma(\mathcal{O})$ die *Borel-σ-Algebra* auf $(\Omega, \mathcal{O})$.

Hat $\mathcal{O}$ eine abzählbare Basis[5] C, so ist $\sigma(\mathcal{O}) = \sigma(C)$. ($\rightarrow$ Übung).

Insbesondere erzeugt $C = \{(-\infty, x] : x \in \mathbb{Q}\}$ die Borel-σ-Algebra $\mathscr{B}(\mathbb{R})$. Das gilt auch für die 2-Punkt-Kompaktifizierung

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}.$$

[5] Eine Menge C heißt Basis des topologischen Raumes $(\Omega, \mathcal{O})$, falls sich jede offene Menge als Vereinigung beliebig vieler Mengen aus C schreiben lässt.

Oft gibt es geeignete, einfache erzeugende Systeme, etwa Monotone Klassen oder Dynkin Systeme, siehe Definitionen A.4 und A.5.

Maße

Nun kommen wir zu dem wichtigen und zentralen Begriff **Maß**. Im Gegensatz zur Analysis sind wir an Maßen interesssiert, die endlich sind, genauer: Wahrscheinlichkeitsmaße sind auf 1 normiert, d.h. $P(\Omega) = 1$. Für allgemeinere Maße nehmen wir oft das Symbol μ.

Definition 4. Sei $\mathscr{F}$ eine σ-Algebra und $\mu : \mathscr{F} \to \mathbb{R}_{\geq 0}$. Gilt für paarweise disjunkte $(A_i) \in \mathscr{F}$

(i) $\mu(\varnothing) = 0$

(ii) $\mu\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

so heißt μ *Maß*.

μ heißt *endlich*, falls $\mu(\Omega) < \infty$ und *σ-finit*, falls es $(A_n)_{n \in \mathbb{N}} \in \mathscr{F}$ gibt, so dass $\sum_{n \in \mathbb{N}} A_n = \Omega$ und $\mu(A_n) < \infty$, für alle $n \in \mathbb{N}$.

Die Eigenschaft (ii) heißt σ-Additivität und ist charakteristisch für ein Maß. Für einen Wahrscheinlichkeitsinhalt lassen sich nur endliche Vereinigungen bilden. Die σ-Additivität eröffnet allerdings die Tür zu Grenzwerten.

Im Anhang werden einige Eigenschaften von Maßen bewiesen, auf die an dieser Stelle verwiesen wird. Wichtig ist zum Beispiel die

Stetigkeit von Maßen, die man unmittelbar aus der σ-Additivität ableiten kann.

Zunächst erhält man leicht (siehe Lemma A.9) dass für die Vereinigung von Mengen $A_1, \dots, A_n$ eine Darstellung über disjunkte Mengen existiert, d.h. es gibt $B_1, \dots, B_n$, paarweise disjunkt, so dass

$$\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} B_i.$$

Hiermit erhält man **Stetigkeit** von unten (bzw. oben) eines Maßes: Sind $A_1 \subseteq A_2 \subseteq \dots$ messbare Mengen, so gilt[6]

[6] Siehe Satz A.14.

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} P(A_i). \tag{5}$$

Maße auf $\mathbb{R}$

Wir können das Lebesgue-Maß λ eindeutig auf $\mathscr{B}(\mathbb{R})$ definieren durch

$$\lambda((a,b]) = b - a, \qquad a, b \in \mathbb{Q} \text{ mit } a \leq b.$$

Die abzählbare Menge der links offenen und rechts abgeschlossenen Intervalle mit rationalen Randpunkten erzeugt die Borel σ-Algebra. Sie bildet einen *Halbring*, siehe Definition A.8. Die eindeutige Fortsetzung des Lebesgue-Maßes von diesem Halbring auf die Borel σ-Algebra kann mit Hilfe von Theorem A.22 erledigt werden. Wie man im Anhang sieht, benötigt dies jedoch einige Arbeit !

Ebenso erhalten wir eine Charakterisierung aller Wahrscheinlichkeitsmaße auf $\mathscr{B}(\mathbb{R})$.

Satz 6. *Eine Funktion $P : \mathscr{B}(\mathbb{R}) \to [0,1]$ ist ein Wahrscheinlichkeitsmaß genau dann, wenn es eine wachsende rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ gibt, so dass*

$$P((a,b]) = F(b) - F(a), \qquad a, b \in \mathbb{Q}, a \leq b.$$

Offensichtlich ist P eindeutig durch F bestimmt! Den Beweis führen wir als Übungsaufgabe.

Für uns wichtig werden Verteilungen, also Bildmaße sein.

Definition 7. Ein Tupel $(\Omega, \mathscr{F})$ heißt *Messraum*, falls $\mathscr{F}$ eine σ-Algebra auf Ω ist. Darüber hinaus heißt das Tripel $(\Omega, \mathscr{F}, \mu)$ heißt *Maßraum*, falls zusätzlich μ Maß auf $(\Omega, \mathscr{F})$ ist.

Definition 8. Sei $(\Omega, \mathscr{F}, \mu)$ ein Maßraum, $\mathscr{F}'$ eine σ-Algebra auf Ω'. Eine Funktion $f : \Omega \to \Omega'$ heißt *$\mathscr{F}$-$\mathscr{F}'$-messbar*, falls

$$f^{-1}(A) \in \mathscr{F} \text{ für alle } A \in \mathscr{F}'.$$

Für ein solches f heißt $f_\# \mu : \mathscr{F}' \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, definiert durch

$$f_\# \mu(A') = \mu(f^{-1}(A')) = \mu(\omega \in \Omega : f(\omega) \in A'), \qquad A' \in \mathscr{F}'$$

Bildmaß von μ unter f.

Im Englischen nennt man das Bildmaß *push-forward*. Sind die σ-Algebren aus dem Kontext klar, so nennt man f einfach meßbar.

Man zeigt leicht, dass $f_\# \mu$ wieder ein Maß ist. Es ist die so genannte *Verteilung* der Zufallsvariablen f. Die **von f erzeugte** σ-Algebra

$$\sigma(f) := f^{-1}(\mathscr{F}')$$

ist in der Tat eine σ-Algebra. Ebenso gilt für $\sigma(\mathscr{C}') = \mathscr{F}'$, dass

$$\sigma(f^{-1}(\mathscr{C}')) = f^{-1}(\sigma(\mathscr{C}')).$$

Ist $(\Omega', \mathscr{F}') = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, so heißt f **reellwertig** und ist f $\mathscr{F}$-$\mathscr{B}(\mathbb{R})$-messbar, so nennen wir f **Borel-messbar**.

Messbarkeit ist ein wichtiger Begriff, so dass wir ein paar Rechenregeln wiederholen. Wir schreiben $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Für den Begriff *Stetigkeit* reicht es, auf topologischen Räumen zu arbeiten: Dort sind Funktionen stetig, wenn die Urbilder offener Mengen wieder offen sind.

Lemma 9. *Seien $(\Omega, \mathscr{F})$, $(\Omega', \mathscr{F}')$, $(\Omega'', \mathscr{F}'')$ Messräume und $f : \Omega \to \Omega'$, sowie $g : \Omega' \to \Omega''$.*

(i) $\mathscr{F}' = \sigma(\mathscr{C}') \Rightarrow f$ *messbar* $\Leftrightarrow f^{-1}(\mathscr{C}') \subseteq \mathscr{F}$

(ii) f, g *messbar* $\Rightarrow f \circ g$ *messbar*

(iii) stetige Abbildungen sind messbar bzgl. der Borel-σ-Algebra.

(iv) Eine reellwertige Funktion f ist genau dann messbar, falls

$$\{\omega : f(\omega) \leq x\} \in \mathscr{F} \qquad \forall x \in \mathbb{Q}$$

(v) $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ *ist messbar falls $A_1, \ldots, A_n \in \mathscr{F}$.*

(vi) $f, g : \Omega \to \overline{\mathbb{R}}$ *messbar* $\Rightarrow f \cdot g,\ a \cdot f + b \cdot g,\ \frac{f}{g} \mathbb{1}_{\{g \neq 0\}}$ *messbar* $(a, b \in \mathbb{R})$

(vii) Sind $f_1, f_2, \cdots : \Omega \to \overline{\mathbb{R}}$ messbar, so auch

$$\sup_n f_n, \inf_n f_n, \limsup_n f_n, \liminf_n f_n.$$

Am interessantesten ist wohl der Beweis für die Messbarkeit von $\sup_n f_n$:

$$\{\omega : \sup f_n(\omega) \leq x\} = \bigcap_{n \geq 1} \underbrace{\{\omega : f_n(\omega) \leq x\}}_{\in \mathscr{F}} \in \mathscr{F}.$$

Approximation von messbaren Funktionen

Eine zentrale Eigenschafte: Meßbare Funktionen sind monoton durch einfache Funktionen approximierbar.

Eine wichtige Eigenschaft messbarer Funktionen ist, dass sie stets durch einfache Funktionen approximierbar sind: Wir betrachten $f \geq 0$ und setzen

$$A_{j,n} = \begin{cases} \{\frac{j}{2^n} < f \leq \frac{j+1}{2^n}\}, & j = 0, \dots, n \cdot 2^n - 1 \\ \{f \geq n\}, & j = n \cdot 2^n. \end{cases}$$

Unsere Apprxoimation ist

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} \mathbb{1}_{\{A_{j,n}\}}.$$

Dann gilt $f_n \uparrow f$. Dies ist der Schlüssel zum Integral! Wir definieren für einfache Funktionen

Natürlich ist
$$\int c\mathbb{1}_A d\mu = c\mu(A).$$
und hieraus entsteht sofort das Integral für einfache Funktionen.

$$\int f d\mu = \int \sum_{i=1}^{n} c_i \mathbb{1}_{A_i} d\mu = \sum_{i=1}^{n} c_i \int_{A_i} d\mu = \sum_{i=1}^{n} c_i \mu(A_i)$$

und für messbare, nicht-negative Funktionen und $f_n \uparrow f$

$$\int f d\mu = \lim_{n \to \infty} \int f_n d\mu = \sup \Big\{ \int g d\mu : g \text{ einfach und } 0 \leq g \leq f \Big\}.$$

Dieses Integral können wir fortsetzten, falls mindestens ein Integral $\int f^+ d\mu, \int f^- d\mu$ endlich ist.

Dieses Prozedere kann man auch auf allgemeinere Situationen übertragen: Das *Bochner*-Integral ist für Banachraum-wertige Zufallszahlen definiert. Es gibt noch weitere Verallgemeinerungen wie das *Pettis*-Integral, wozu ein schwacher Messbarkeitsbegriff eingeführt wird. Erste Definitionen und Beispiele finden sich beispielsweise auf Wikipedia.

Dann definieren wir

$$\mathcal{L}^1(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{Q}} : \int |f| d\mu < \infty \Big\} \qquad \text{(und analog } \mathcal{L}^p)$$

Leider hat dieser Raum schlechte Trennungseigenschaften: Gibt es eine nichtleere Null-Menge, so gibt es immer verschiedene meßbare

Funktionen, die den Abstand Null haben. Um das zu beheben betrachtet man Äquivalenzklassen, siehe auch Abschnitt VI.2 in Elstrodt (2013) [7].

Auf den messbaren Funktionen kann man folgende Äquivalenzrelation einführen: $f \sim g$ falls $P(f = g) = 1$. Die hierdurch erhaltenen Äquivalenzklassen definieren die L^p-Räume. In der Notation machen wir dies durch die Unterscheidung $\mathcal{L}$ und L kenntlich. Im Folgenden kürzt f. s. fast sicher ab, etwa $f \leq g$ f. s. ist gleichbedeutent mit $\mu(\omega \in \Omega : f > g) = 0$. Meist ist dies auch aus dem Kontext klar und wir lassen das f.s. weg.

[7] Jürgen Elstrodt. *Maß-und Integrationstheorie*. Springer, 2018

Die L^p-Räume sind die Äquivalenzklassen der integrierbaren Funktionen. Für $p \geq 1$ erhält man einen metrischen Raum der vollständig ist (Banachraum), mit $p = 2$ sogar einen Hilbertraum.

Satz 10. *Erste Eigenschaften:*

(i) $f \leq g$ *(f. s.)* $\Rightarrow \int f d\mu \leq \int g d\mu$

(ii) $|\int f d\mu| \leq \int |f| d\mu$ *(Dreiecksungleichung)*

(iii) $\int (af + bg) d\mu = \int f d\mu + b \int g d\mu$ *(Linearität)*

(iv) $f = 0$ *(f. s.)* $\Rightarrow \int f d\mu = 0$

(v) $\int f d\mu < \infty \Rightarrow f < \infty$ *(f.s.)*

(vi) $0 \leq f_n \uparrow f \Rightarrow \int f_n d\mu \uparrow \int f d\mu.$ *(Monotone Konvergenz)*

Diese Eigenschaften erhält man recht direkt aus der Definition des Integrals. Als Übung beweisen wir den Substitutionssatz

Satz 11. $g \in \mathcal{L}^1(f_\# \mu) \Rightarrow g \circ f \in \mathcal{L}^1(\mu)$ *und*

$$\int g \circ f \, d\mu = \int g \, d(f_\# \mu).$$

Beweis. Es genügt, die Aussage für einfache nicht-negative g zu zeigen. Der allgemeine Fall folgt durch Approximation.

Sei $g = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \Rightarrow g \circ f = \sum_{i=1}^n c_i \mathbb{1}_{\{f \in A_i\}}$, also

$$\int g \circ f d\mu = \sum_{i=1}^n c_i \, \mu(f \in A_i) = \sum_{i=12}^n c_i \, f_\# \mu(A_i) = \int g \, d(f_\# \mu). \qquad \square$$

So wird man viele Aussage für Integrale beweisen können: Zunächst die Aussagen auf einfachen, nicht negativen Funktionen zeigen und dann durch Approximation das Resultat auf die Integrale übertragen.

Konvergenzsätze für das Integral

Äußerst wichtig sind die folgenden Konvergenzsätze. Sie sind ein unerlässliches Hilfsmittel in der Wahrscheinlichkeitstheorie.

Theorem 12 (Monotone Konvergenz). *Sei* $(\Omega, \mathscr{F}, \mu)$ *Maßraum,* $f_n \geq 0$ *und* $f : \Omega \to \overline{\mathbb{R}}$ *messbar mit* $P(f_n \uparrow f) = 1 \Rightarrow$

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Beweis. Folgt direkt aus der monotonen Konvergenz (Teil (vi) in Satz 10) mit

$$g_n = (f - f_n)\mathbb{1}_{\{\omega\in\Omega: f_n(\omega)\uparrow f(\omega)\}}.$$ □

Theorem 13 (Fatou). *Sei* $(\Omega, \mathscr{F}, \mu)$ *Maßraum* $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$ nichtnegativ *und messbar. Dann gilt*

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \liminf_{n\to\infty} f_n d\mu.$$

Das Lemma von Fatou kommt mit den geringsten Voraussetzungen aus (lediglich nicht-negativität, also eine untere Schranke wird gefordert). Allerdings erhält man auch nur eine schwächere Aussage.

Beweis. $k \geq n \Rightarrow f_k \geq \inf_{m\geq n} f_m$, so dass

$$\inf_{k\geq n} \int f_k d\mu \geq \int \inf_{k\geq n} f_n d\mu.$$

Für $n \to \infty$ erhalten wir

$$\liminf_{n\to\infty} \int f_n d\mu \geq \sup_{n\in\mathbb{N}} \int \inf_{k\geq n} f_k d\mu = \int \liminf_{n\to\infty} f_n d\mu$$

wegen monotoner Konvergenz $(\inf_{k\geq n} f_k \uparrow \liminf_{n\to\infty} f_n)$. □

Theorem 14 (Majorisierte Konvergenz). *Sei* $(\Omega, \mathscr{F}, \mu)$ *Maßraum und* $f, g, (f_n)_{n\geq 1} : \Omega \to \overline{\mathbb{R}}$ *messbar. Sei* $|f_n| \leq g$ *f. s. und* $\lim_{n\to\infty} f_n = f$, $g \in \mathcal{L}^1(\mu)$.

Dann gilt

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Die Beweisidee ist Fatou auf $g + f$, $g - f$ anzuwenden.

Beweis. Ohne Einschränkung sei $|f_n| \leq g \ \forall \omega \in \Omega \Rightarrow$. Zunächst erhalten wir mit Fatou, dass $(g + f_n \geq 0)$

$$\int (g+f)d\mu \leq \liminf_{n\to\infty} \int (g+f_n)d\mu = \int g d\mu + \liminf_{n\to\infty} \int f_n d\mu \tag{1}$$

Umgekehrt erhalten wir

$$\int (g-f)d\mu \leq \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu. \tag{2}$$

Beides zusammen liefert

$$\int f d\mu \overset{(1)}{\leq} \liminf_{n\to\infty} \int f_n d\mu \leq \limsup_{n\to\infty} \int f_n d\mu \overset{(2)}{\leq} \int f d\mu.$$ □

Die L^p-Räume

Nun lernen wir eine wichtige Klasse von Metriken und damit Topologien für Zufallsvariablen kennen: Die L^p-Räume. Sie sind auch ein wichtiges Beispiel in der Funktionalanalysis, insbesondere da man die Dualräume sehr explizit beschreiben kann. Für uns werden sie vor allem einen ersten Konvergenzbegriff von Zufallsvariablen erzeugen. Um den Funktionalanalytischen Charakter zu betonen sprechen wir zunächst von (messbaren) Funktionen - das sind aber natürlich später unsere Zufallsvariablen.

Wir definieren für $0 < p < \infty$ die Halbnorm

$$\|f\|_p = \Big(\int |f|^p d\mu\Big)^{1/p}$$

auf dem Raum der p-fach integrierbaren Funktionen

$$\mathcal{L}^p(\mu) = \{f : \Omega \to \overline{\mathbb{R}} \text{ messbar mit } \|f\|_p < \infty\}$$

Dieser Raum ist vollständig, aber da es mehrere Funktionen mit $\|f\|_p = 0$ gibt, sind Limiten nicht eindeutig bestimmt. Man geht deswegen zu dem Quotientenraum[8]

$$L^p(\mu) := \mathcal{L}^p(\mu)/N_p$$

mit $N_p = \{f : \|f\|_p = 0\}$ über. Dieser ist ein Banachraum für $p \geq 1$.

Für $p = \infty$ erhält man in gewissem Sinne Funktionen die bezüglich jeder L^p-norm integrierbar sind - dass sind beschränkte Funktionen. So definiert man den Raum der (fast sicher) beschränkte Funktionen $L^\infty(\mu)$ durch die Norm

$$\|f\|_\infty = \inf\{K : \mu(|f| > K) = 0\}$$

(ebenfalls ein Banachraum).

Ähnlich geht man in die andere Richtung vor: Für $p \to 0$ fallen alle Beschränkungen weg und man definiert den Raum $L^0(\mu)$ als den Raum aller meßbaren Funktionen.

Unter anderem gilt für $f \in \mathcal{L}^p(\mu)$ und $g \in \mathcal{L}^q(\mu)$, dass[9]

(i) $0 < p, q, r \leq \infty$ und $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$: $\quad \|fg\|_r = \|f\|_p \|g\|_q$

(ii) $1 \leq p \leq \infty$: $\quad \|f + g\|_p \leq \|f\|_p + \|g\|_p$

Für weitere Details zu den L^p-Räumen siehe etwa Werner[10], Kapitel 1.1 Beispiel (h). Auf den L^p-Räumen haben wir einen Abstand, über welchen wir ganz natürlich einen ersten Konvergenzbegriff für Zufallsvariablen einführen können.

[8] Das macht in der Regel keine Probleme. Allerdings ist etwa die Abbildung $f \mapsto f(x_0)$ nicht mehr wohldefiniert.

[9] Die Hölder- und Dreiecksungleichung.

[10] D. Werner. *Funktionalanalysis*. Springer, 2000

Definition 15 (Konvergenz im p-ten Mittel)**.** Eine Folge $(f_n)_{n\geq 1} \subset L^p(\mu)$ **konvergiert im p-ten Mittel**, falls

$$\|f_n - f\|_p \underset{n\to\infty}{\longrightarrow} 0.$$

Wir schreiben

$$f_n \xrightarrow[n\to\infty]{L^p(\mu)} f.$$

Konvergiert eine Folge im p-ten Mittel, so auch im q-ten Mittel, falls $q < p$. (Hölder)

Wie bereits erwähnt, ist $L^p(\mu)$, $p \geq 1$ **vollständig** (jede Cauchy-Folge konvergiert) und somit ein Banachraum.

Für die Behandlung des Hilbertraums L^2 sei auf Kapitel im Anhang verwiesen.

Wahrscheinlichkeitstheorie

Nach dieser Einführung in die technischen Hilfsmittel, Maß und Integral, kommen wir nun zum Kern dieser Vorlesung: Zufallsvarblen, deren Charakteristika und zugehörige Methoden wie Martingale und Grenzwertsätze.

Grundlagen

Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ und einen Bildraum $(\Omega', \mathscr{F}')$. Eine Abbildung

$$X : \Omega \to \Omega'$$

heißt **Zufallsvariable**, falls sie $\mathscr{F} - \mathscr{F}'$-messbar ist. Wir schreiben

$$\sigma(X) := \{X^{-1}(B) : B \in \mathscr{F}'\},$$

mit $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$ für die von X erzeugte σ-Algebra.

Das *Bildmaß* $X_\# P$ gegeben durch[11]

$$X_\# P(B) = P(X^{-1}(B)) : \mathscr{F}' \to [0,1]$$

heißt **Verteilung** von X. Haben X und Y die gleiche Verteilung (are equal in law), so schreiben wir

$$X \overset{\mathscr{L}}{=} Y.$$

Ein zentrales Hilfsmittel zur Beschreibung einer Zufallsvariable ist die Dichte: Wir sagen X hat die *Dichte* f bzgl. des Maßes ν, falls

$$P(X \in A) = \int_A f d\nu, \qquad A \in \mathscr{F}' \tag{1}$$

gilt. In unserer Notation mit dem Push-Forward ist das genau dann der Fall, falls $X_\# P = f \cdot \nu$.

[11] Vergleiche Definition A.25.

Der Erwartungswert. Der **Erwartungswert** einer Zufallsvariablen ist definiert durch

$$E[X] := \int X \, dP, \tag{2}$$

falls $E[|X|] < \infty$. Wir schreiben dann auch $X \in \mathcal{L}^1(P)$ (bzw. $X \in L^1(P)$ oder schlicht – falls es klar ist um welches Maß es sich handelt - $X \in L^1$).

Ist zumindest eines der Integrale $E[X^+]$ und $E[X^-]$ endlich, so nennt man X *quasi-integrierbar* und definiert als Konvention $E[X] = +\infty$, womit dann auch der Erwartungswert wohldefiniert ist. Ist keins der beiden Integrale endlich (zum Beispiel bei einer Cauchy-Verteilung), so kann man mit dem Erwartungswert (und allen höheren Momenten) nicht viel anfangen.

Momente. Das k-te **Moment** der Zufallsvariablen X ist definiert durch

$$m_X(k) := E[X^k], \tag{3}$$

sofern das Integral wohldefiniert ist. Das **Momentenproblem** ist die Frage inwieweit die Momente eine Verteilung genau festlegen: Kenne ich alle Momente $m_X(k),\ k = 1, 2, \ldots$ - ist die Verteilung genau bestimmt . Die Antwort ist: Unter geeigneten Voraussetzungen, ja! Siehe [12], Korollar 15.32.

siehe etwa `spektrum.de` - Lexikon der Mathematik - Momentenproblem

[12] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2013

Monotonie und Linearität des Integrals liefern für Zufallsvariable X, Y die wichtige Regel

$$E[aX + bY] = aE[X] + bE[Y].$$

Außerdem gilt, dass $0 \le X \le Y \Rightarrow E[X] \le E[Y]$.

Beispiele

Betrachten wir ein paar Beispiele - diese kennen wir schon aus der Stochastik 1. Diese und viele weitere (und deren Anwendungen) findet man z.B. auch in [13]. Wir betrachten Zufallsvariablen mit Dichte f bezüglich des Lesbesguemaßes, siehe Gleichung 1.

[13] C. Czado and T. Schmidt. *Mathematische Statistik*. Springer Verlag. Berlin Heidelberg New York, 2011

- *Exponentialverteilung:* Wir schreiben $X \sim \text{Exp}(\lambda)$, falls $\lambda > 0$ und

$$f(x) = \mathbb{1}_{\{x>0\}} \lambda e^{-\lambda x}.$$

- *Gleichverteilung:* Wir schreiben $X \sim U(a, b)$, falls $a < b$ und

$$f(x) = \mathbb{1}_{\{x \in [a,b]\}} \frac{1}{b-a}.$$

- *Normalverteilung:* Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$, falls $\mu \in \mathbb{R}$, $\sigma > 0$ und
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4}$$
Dann gilt, dass $E[X] = \mu$ und $\operatorname{Var}(X) = \sigma^2$. Ist $\mu = 0$ und $\sigma = 1$, so spricht man von einer *Standardnormalverteilung*.

Die von X erzeugte σ-Algebra

Die von X erzeugte σ-Algebra hat folgende besonders weitreichende Eigenschaft:

Lemma 5. *Sei $X : \Omega \to \Omega'$ eine Zufallsvariable. $Z : \Omega \to \overline{\mathbb{R}}$ ist genau dann $\sigma(X)$-messbar, falls*
$$Z = f(X)$$
mit $f : \Omega' \to \overline{\mathbb{R}}$ $\mathscr{F}' - (\overline{\mathbb{R}})$-messbar.

Beweis. „$\Leftarrow$" klar.

„$\Rightarrow$" Ohne Beschränkung der Allgemeinheit[14] sei $Z \geq 0$.

(i) Sei $Z = \mathbb{1}_A$, $A \in \sigma(X) \Rightarrow \exists A' \in \mathscr{F}'$ mit $X^{-1}(A') = A$, also
$$Z = \mathbb{1}_A = \mathbb{1}_{X^{-1}(A')} = \mathbb{1}_{A'} \circ X,$$
wir können also $f = \mathbb{1}_{A'}$ wählen.

(ii) Mit Linearität erhalten wir die Aussage für einfache Z

(iii) Ist $Z \geq 0$ messbar, so gibt es einfache $(Z^n)_{n\geq 1} \uparrow Z$ und $Z_n = f_n \circ X \;\Rightarrow$
$$(\sup_n f_n) \circ X = \sup_n f_n \circ X = \sup Z_n = Z$$
und mit $f = \sup f_n$ sind wir fertig. $\square$

[14] Ansonsten betrachte $Z = Z^+ - Z^-$.

Dieser Satz ist ein weiteres schönes Beispiel für die bereits erwähnte Beweistechnik, maßtheoretische Resultate zunächst für Indikatoren zu beweisen und dann messbare Funktionen durch einfache Funktionen monoton zu approximieren.

Wichtige Ungleichungen

Wir lernen nun einige wichtige Ungleichungen kennen.

Satz 6. *Seien X, Y reellwertige Zufallsvariable. Dann gilt*

(i) für alle monoton wachsenden $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ und $\varepsilon > 0$ so dass $f(\varepsilon) > 0$

$$P(|X| \geq \varepsilon) \leq \frac{E[f(|X|)]}{f(\varepsilon)}$$

(ii) für $E[X^2] < \infty$, dass

$$P(|X - E[X]| > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

(iii) $0 < p, q, r \leq \infty$ mit $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$

$$\|XY\|_r \leq \|X\|_p \, \|Y\|_q \tag{7}$$

(iv) $1 \leq p \leq \infty$:

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

(v) $0 < p < 1$:

$$E[|X + Y|^p] \leq E[|X|^p] + E[|Y|^p]$$

(vi) $0 < p \leq q$, $X \in \mathcal{L}^q \Rightarrow$

$$\|X\|_p \leq \|X\|_q$$

Markov-Ungleichung

Tschebyscheff-Ungleichung

Hölder-Ungleichung
p=q=2: Cauchy-Schwarz-Ugl.

Minkowski-Ugl.

Beweis. (i) Da $f(\varepsilon) > 0$ und $f(x) \geq 0$ folgt für alle $x \in \mathbb{R}_{\geq 0}$, dass

$$\begin{aligned}
\mathbb{1}_{\{|x| \geq \varepsilon\}} \leq \mathbb{1}_{\{f(|x|) \geq f(\varepsilon)\}} &= \mathbb{1}_{\left\{\frac{f(|x|)}{f(\varepsilon)} \geq 1\right\}} \\
&\leq \frac{f(|x|)}{f(\varepsilon)} \mathbb{1}_{\left\{\frac{f(|x|)}{f(\varepsilon)} \geq \mathbb{1}\right\}} \\
&\leq \frac{f(|x|)}{f(\varepsilon)}
\end{aligned}$$

Durch Monotonie des Erwartungswertes folgt die Behauptung.

(ii) Wir verwenden $Y = X - E[X]$ und $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} : x \mapsto x^2$ (monoton auf $\mathbb{R}_{\geq 0}$) in (i)

(iii) Wir nutzen Konvexität der exponential Funktion: Da e^x konvex ist, gilt für $x, y \geq 0$

$$(xy)^r = \exp(r \log x + r \log y) = \exp\left(\frac{r}{p} \log x^p + \frac{r}{q} \log x^q\right)$$
$$\leq \frac{r}{p} x^p + \frac{r}{q} x^p \qquad (\text{da } \tfrac{r}{p} + \tfrac{r}{q} = 1)$$

Das wenden wir an auf $X' = \frac{|X|}{\|X\|_p}$, $Y' = \frac{|Y|}{\|Y\|_p}$ und erhalten

$$E[(X'Y')^r] \leq \frac{r}{p} \frac{E[|X|^p]}{\|X\|_p^p} + \frac{r}{q} \frac{E[|Y|^q]}{\|Y\|_q^q} = \frac{r}{p} + \frac{r}{q} = \frac{r}{r} = 1,$$

also

$$1 \geq E[(X'Y')^r] = \frac{E[|X'Y'|^r]}{\|X\|_p^r \|Y\|_q^r}$$

und die Behauptung folgt.

(iv) Dieser Beweis ist trickreicher. Zunächst ist

$$|x+y|^p = |x+y| \cdot |x+y|^{p-1} \leq |x| \cdot |x+y|^{p-1} + |y| \cdot |x+y|^{p-1}$$

und mit (iii)

$$E[|X| \cdot |X+Y|^{p-1}] \leq \|X\|_p \cdot \||X+Y|^{p-1}\|_q$$

da mit $\frac{1}{p} + \frac{1}{q} = 1$. Hieraus folgt auch, dass $(p-1) \cdot q = p$ und $\frac{p}{q} = p - 1$. Damit ist

$$\||X+Y|^{p-1}\|_q = E\left[|X+Y|^{(p-1)\cdot q}\right]^{1/q} = \|X+Y\|_p^{p/q} = \|X+Y\|_p^{p-1}.$$

Das liefert

$$E[|X+Y|^p] = \|X+Y\|_p^p \leq \left(\|X\|_p + \|Y\|_p\right) \cdot \|X+Y\|_p^{p-1}.$$

(v) Wir halten y fest und betrachten die Funktion

$$f(x) = (x+y)^p - x^p - y^p.$$

Dann ist zunächst einmal $f(0) = 0$ (für alle y). Weiterhin ist die Ableitung

$$f'(x) = p(x+y)^{p-1} - px^{p-1} \leq 0,$$

da $x + y \geq y$ und die Abbildung $x \mapsto x^q$ ist fallend für $q = p - 1 < 0$. Damit ist also $f(x) \leq f(0)$ und somit folgt die Behauptung.

(vi) Zunächst ist $x \mapsto x^{p/q}$ konkav auf $\mathbb{R}_{\geq 0}$, also folgt mit Satz 8

$$E[|X|^p] = E[(|X|^q)^{p/q}] \leq (E[|X|^q])^{p/q}$$

und wir sind fertig.

$\square$

Satz 8 (Jensensche Ungleichung). *Sei $g : \mathbb{R} \to \mathbb{R}$ konvex und $X \in \mathcal{L}^1$. Dann gilt*

$$E[g(X)] \geq g(E[X]).$$

Gleichheit gilt genau dann, wenn für jede Gerade $a + bx$ tangential zu g an $x = E[X]$ gilt, dass $P(g(X) = a + bX) = 1$.

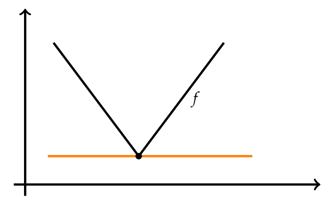
Beweis. Für eine konvexe Funktion f gibt es an jedem Punkt x eine Tangente $y \mapsto a + by$, so dass

$$f(x) \geq a + bx \quad \forall x.$$

Wir betrachten diese unterstützende Gerade am Punkt $E[X]$. Dann ist

$$\begin{aligned} f(x) &\geq f(E[X]) + b \cdot (x - E[X]), \quad \text{also} \\ E[f(x)] &\geq f(E[X]) + b \cdot (E[x] - E[X]) = f(E[X]) \end{aligned}$$

und die Behauptung folgt. □



Für eine konvexe Funktion existiert an jedem Punkt eine Tangente, die von der Funktion selbst dominiert wird.

Charakteristische Funktionen

Als ein wichtiges Hilfsmittel zur Charakterisierung von Verteilungen erweisen sich Fourier- und Laplace-Transformierte. Damit spielen sie eine wichtige Rolle z.B. für den zentralen Grenzwertsatz.

Definition 9. Ist X eine d-dimensionale Zufallsvariable, so heißt $\varphi : \mathbb{R}^d \to \mathbb{C}$, gegeben durch

$$\varphi(t) = E[e^{i\langle t,x\rangle}]$$

charakteristische Funktion oder *Fourier-Transformierte* von X.

Die Funktion $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, gegeben durch

$$\mathcal{L}(t) = E[e^{-\langle t,x\rangle}]$$

heißt *Laplace-Transformierte* von X.

Satz 10. *Es gilt*

(i) $|\varphi(t)| \leq 1$ *für alle* $t \in \mathbb{R}^d$ *und* $\varphi(0) = 1$.

(ii) φ *ist gleichmäßig stetig,*

(iii) $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at) \qquad a \in \mathbb{R}, b \in \mathbb{R}^d, t \in \mathbb{R}^d$

Beweis. (i) und (iii) sind klar.

Schreiben wir kurz $tX = \langle t, X \rangle$. Für (ii) verwenden wir, dass mit majorisierter Konvergenz folgt:

$$\left|E[e^{itX} \cdot (e^{ihX} - 1)]\right| \overset{(i)}{\leq} \left|E[(e^{ihX} - 1)]\right| \underset{h \to 0}{\longrightarrow} 0,$$

also $\sup_{t \in \mathbb{R}} |\varphi(t+h) - \varphi(t)| \leq |E[(e^{ihX} - 1)]| \underset{h \to 0}{\longrightarrow} 0$, so dass φ gleich-mäßig stetig ist. □

Die charakteristische Funktion bestimmt das Maß P genau, wie folgender Satz zeigt. Der Beweis ist maßtheoretischer Natur, basiert aber darauf, dass wir für eine abgeschlossene Menge eine sogar Lipschitzstetige Funktion finden können, die den Indikator $\mathbb{1}_A$ approximiert. Für ein Intervall $[a, b]$ kann man sich das sehr gut vorstellen. Wir verweisen auf Lemma 13.10 und Satz 13.11 in Klenke[15]. Damit erhalten wir folgenden Satz (Satz 15.8 in Klenke):

[15] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2013

Satz 11. *Ein endliches Maß μ auf $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ ist durch seine charakteristische Funktion eindeutig festgelegt.*

Beispiel 12. Folgende Beispiele sind wichtige Fourier-Transformierte:

(i) Für $X \sim B(n, p)$ ist

$$E[e^{itX}] = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-h} e^{itk} = \left(1 - p + pe^{it}\right)^n$$

(ii) Für $X \sim \mathrm{Poiss}(\lambda)$ ist

$$E[e^{itX}] = \sum_{k \geq 0} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}$$

(iii) Für $X \sim \mathcal{N}(\mu, \sigma^2)$ ist zunächst

$$E[e^{itX}] = E[e^{it\mu + it\sigma\xi}]$$

mit $\xi \sim \mathcal{N}(0, 1)$. Weiterhin ist

$$E[e^{it\xi}] = \frac{1}{\sqrt{2\pi}} \int e^{itx} e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}$$

und somit

$$\varphi(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

(iv) Ist $X \sim \mathrm{Exp}(\lambda)$, so folgt

$$E[e^{-tX}] = \int_0^\infty e^{-tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda + t}.$$

Satz 13. *Sei X eine reellwertige Zufallsvariable.*

(i) *Ist $X \in L^p$, so ist φ_X p-mal stetig differenzierbar und*

$$\varphi_X^{(k)}(t) = E[(iX)^k e^{itX}], \qquad k = 0, \dots, p$$

(ii) *Für $X \in L^2$ gilt*

$$\varphi(t) = 1 + itE[X] - \frac{t^2}{2} E[X^2] + \varepsilon(t) \cdot t^2$$

mit $\varepsilon(t) \xrightarrow[t \to 0]{} 0$.

Beweis. (i) Da $X \in L^p$, existiert der Erwartungswert für $k \leq p$. Wir nutzen Induktion: $k = 0$ klar. Gelte die Behauptung für ein k. Dann ist

$$\left| \frac{d^{k+1} e^{itx}}{dt^{k+1}} \right| = \lim_{h \to 0} \left| \frac{(ix)^k e^{i(t+h)x} - (ix)^k e^{itx}}{h} \right| \leq \lim_{h \to 0} \left| x^k \frac{e^{ihx} - 1}{h} \right|$$

Nun ist

$$\begin{aligned} \lim_{h \to 0} \frac{e^{ihx} - 1}{h} &= \lim_{n \to 0} \left| \frac{xih + \frac{(ihx)^2}{2} + \cdots}{h} \right| \\ &\leq |x| \cdot \lim_{h \to 0} \left(1 + |hx| + \frac{|hx|^2}{2} + \cdots \right) \leq |x|, \end{aligned}$$

Und somit gibt es für $\frac{d^{k+1} e^{itx}}{dt^{k+1}}$ eine integrierbare Majorante. Damit folgt

$$\varphi^{(kn)}(t) = E\Big[\frac{d}{dt} (iX)^k e^{itX} \Big] = E[(iX)^{k+1} e^{itX}].$$

Ebenso folgert man Stetigkeit der Ableitung mit majorisierter Konvergenz.

(ii) Die Taylor-Entwicklung mit Restglied liefert

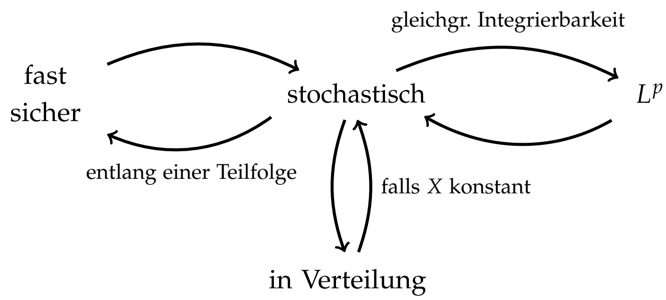
$$e^{itX} = 1 + itX - \frac{t^2 X^2}{2} (\cos \theta_1 tX + i \sin \theta_2 tX)$$

mit Zufallsvarialben θ_1, θ_2 und $|\theta_i| \leq 1$, $i = 1, 2$. Das ergibt

$$\varphi(t) = 1 + itE[X] - \frac{t^2}{2} E[X^2] + \varepsilon |t| t^2$$

mit $2\varepsilon(t) = E[X^2(1 - \cos \theta_1 tX - i \sin \theta_2 tX)] \xrightarrow{t \to 0} 0$ wegen majorisierter Konvergenz. □

Konvergenzen



Wir werden nun verschiedene Konvergenzen kennenlernen: Die **fast sichere Konvergenz**,

$$X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$$

die **stochastische Konvergenz**

$$X_n \xrightarrow[n\to\infty]{P} X$$

und die **Konvergenz bezüglich der $L^p(P)$-Norm**,

$$X_n \xrightarrow[n\to\infty]{L^p} X.$$

Etwas später lernen wir die **schwache Konvergenz** (auch Konvergenz in Verteilung)

$$X_n \xrightarrow[n\to\infty]{\mathscr{L}} X$$

kennen.

Hierbei ist fast sichere Konvergenz der stärkste Begriff - Konvergenz im klassischen Sinne für fast alle $\omega \in \Omega$. Die stochastische Konvergenz ist deutlich schwächer und oft leichter zu erreichen. Entlang einer Teilfolge erhalten wir wieder fast sichere Konvergenz.

Die L^p-Konvergenz impliziert die stochastische Konvergenz, auch sie ist stärker als die stochastische Konvergenz, fordert allerdings Existenz von p-ten Momenten. Unter gleichgradiger (uniformer) Integrierbarkeit folgt sie aus der stochastischen und damit auch aus der fast sicheren Konvergenz.

Die schwache Konvergenz, die wir später kennenlernen ist etwas anders geartet. Sie ist gar nicht auf einem festen Wahrscheinlichkeitstraum definiert, folgt aber immer aus stochastischer Konvergenz falls der Grenzwert konstant ist. Umgekehrt impliziert sie stochastische Konvergenz.

In der folgenden Definition führen wir die Begriffe ein.

Definition 14. Seien $X, X_1, X_2, \ldots$ Zufallsvariablen in einem separablen metrischen Raum (E, d).

(i) Gilt

$$\lim_{n\to\infty} d(X_n, X) = 0 \qquad P\text{-f.s.},$$

so konvergiert (X_n) *fast sicher* gegen X

$$(X_n \xrightarrow[n\to\infty]{\text{f.s.}} X).$$

(ii) Gilt für alle $\varepsilon > 0$, dass

$$\lim_{n\to\infty} P(d(X_n, X) > \varepsilon) = 0,$$

so konvergiert (X_n) *stochastisch* gegen X

$$(X_n \xrightarrow[n\to\infty]{P} X).$$

(iii) Sind $X, (X_n)$ reellwertig und gilt für $p > 0$, dass

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0,$$

so konvergiert (X_n) *in* L^p (oder im p-ten Mittel) gegen X

$$(X_n \xrightarrow[n\to\infty]{L_p} X).$$

ÜA: In einem separablen metrischen Raum ist für zwei ZV X und Y die Abbildung $d(X, Y)$ messbar.

Die wichtigste Konvergenzart ist hierbei die fast sichere Konvergenz. Sie ist allerdings oft nur schwer nachzuweisen.

Stochastische Konvergenz impliziert nicht fast sichere Konvergenz

Beispiel 15 (Beispiele). (i) Wir beginnen mit einem Beispiel zur stochastischen Konvergenz Sei

$$A_1 = \left[0, \frac{1}{2}\right], \qquad A_2 = \left[\frac{1}{2}, 1\right],$$
$$A_3 = \left[0, \frac{1}{4}\right], \left[\frac{1}{4}, \frac{2}{4}\right], \ldots, \quad A_6 = \left[\frac{3}{4}, 1\right], \ldots$$

und $X_n = \mathbb{1}_{\{U \in A_n\}}$ mit $U \sim U(0,1)$. Dann ist

$$\lim_{n\to\infty} P(|X_n| > \varepsilon) = \lim_{n\to\infty} P(U \in A_n) = 0,$$

aber $X_n \overset{\text{f.s.}}{\not\longrightarrow} 0$. Genauer: X_n konvergiert an **keinem** ω, denn für jedes $\omega \in \Omega$ und $n \in \mathbb{N}$ gibt es $m > n$ mit $X_m(\omega) = 1$.

(ii) Sei $U \sim U[0,1]$ und $B_n = [0, \frac{1}{n}]$ und $X_n = n \cdot \mathbb{1}_{\{U \in B_n\}}$. Dann $X_n \xrightarrow{\text{f.s.}} 0$, aber

$$E[X_n] = n \cdot P(U \in B_n) = 1.$$

Fast sichere Konvergenz impliziert nicht L^1-Konvergenz

Lemma 16. *Seien $X, Y, X_1, X_2, \dots$ Zufallsvariable auf dem metrischen Raum (E, d). Gilt $X_n \xrightarrow{P} X, X_n \xrightarrow{P} Y$, so folgt*

$$X = Y \ P\text{-f.s.}$$

Der Limes ist eindeutig bestimmt.

Beweis. Aus der stochastischen Konvergenz folgt für alle $\varepsilon > 0$, dass

$$P(d(X,Y) > 2\varepsilon) \leq P\big(d(X_n, X) > \varepsilon \text{ oder } d(X_n, Y) > \varepsilon\big) \underset{n\to\infty}{\rightarrow} 0,$$

also ist $P(d(X,Y) > \varepsilon) = 0$ für alle $\varepsilon > 0$. Hieraus folgt

$$P(X \neq Y) = P(d(X,Y) > 0) = 0$$

mit Hilfe von Satz A.14 (Stetigkeit von oben). □

Das folgende Lemma zeigt, dass stochastische Konvergenz metrisierbar ist, es also eine Metrik und eine zugehörige Topologie gibt, die Topologie zur Konvergenz in Wahrscheinlichkeit (oder allgemeiner bezüglich eines Maßes).

Lemma 17. *Seien $X, X_1, \dots$ Zufallsvariablen auf dem metrischen Raum (E, d). Dann gilt*

$$X_n \xrightarrow[n\to\infty]{P} X \iff E[d(X_n, X) \wedge 1] \xrightarrow[n\to\infty]{} 0.$$

Stochastische Konvergenz ist metrisierbar.

Beweis. „⇒“ Aus $X_n \xrightarrow{P} X$ folgt für alle $\varepsilon > 0$, dass

$$\lim_{n\to\infty} E[d(X_n, X) \wedge 1] \leq \lim_{n\to\infty} (\varepsilon + P(d(X_n, X) > \varepsilon)) = \varepsilon.$$

„⇐“ Umgekehrt folgt mit $0 < \varepsilon \leq 1$ aus der Markov-Ungleichung, dass

$$P(d(X_n, X) > \varepsilon) \leq \frac{E[d(X_n, X) \wedge 1]}{\varepsilon} \underset{n\to\infty}{\rightarrow} 0.$$ □

Satz 18. *Es gelten folgende Aussagen:*

(i) *Fast-sichere impliziert stochastische Konvergenz*

(ii) $\sum_n P(|Z_n - Z| > \varepsilon) < \infty \ \forall \varepsilon > 0 \quad \Rightarrow \quad Z_n \to Z$ *fast sicher.*

(iii) *Gilt* $Z_n \xrightarrow{P} Z$, *so existiert eine Teilfolge* $Z_{n(j)}$, *welche fast sicher gegen* Z *konvergiert.*

Beweis. Teil (i): Zunächst einmal ist

$$1 = P(\lim_n Z_n = Z) = P\Big(\bigcap_{\varepsilon>0} \bigcup_{n\geq 1} \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big). \tag{19}$$

Die Mengen $\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}$ sind monoton wachsend in n und ε, was wir nun geschickt ausnutzen. Mit der σ-Stetigkeit von P folgt, dass (19) äquivalent dazu ist, dass

$$1 = P\Big(\bigcup_{n\geq 1} \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) \tag{20}$$

für alle[16] $\varepsilon > 0$. Eine zweite Anwendung der σ-Stetigkeit liefert, dass (19) äquivalent dazu ist, dass

$$\lim_n P\Big(\bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) = 1 \ \ \forall \varepsilon > 0.$$

Hieraus folgt

$$0 = \lim_n P\Big(\bigcup_{k\geq n} \{|Z_k - Z| > \varepsilon\}\Big) \geq \lim_n P(|Z_n - Z| > \varepsilon)$$

und somit die gesuchte stochastische Konvergenz.

Teil (ii): Für ein festes $\varepsilon > 0$ folgt mit Borel-Cantelli, dass

$$\sum_{n=1}^{\infty} P(|Z_n - Z| > \varepsilon) < 0 \quad \Rightarrow \quad P\Big(\bigcap_n \bigcup_{k\geq n} \{|Z_k - Z| > \varepsilon\}\Big) = 0$$

Das Komplement ist

$$P\Big(\bigcup_n \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) = 1. \tag{21}$$

Da ε beliebig gewählt war, liefert die Äquivalenz (20) nun die fast sichere Konvergenz.

Für Teil (iii) wählen wir $n(j)$ so dass $P\big(|Z_{n(j)} - Z| \geq \frac{1}{j}\big) \leq \frac{1}{j^2}$.

□

[16] In der Tat: Ist $1 = P(\cap_{\varepsilon>0} A_\varepsilon)$ und gilt $A_\varepsilon \supset A_{\varepsilon'}$ für $\varepsilon > \varepsilon'$, so folgt $1 = P(\cap_{n\geq 0} A_{n^{-1}}) = \lim_n P(A_{n^{-1}})$. Da $\lim_n P(A_{n^{-1}}) \leq P(A_{n'^{-1}})$ für alle $n' \geq 1$, folgt sogar, dass $1 = P(A_{n^{-1}})$ für alle $n \geq 1$ und schließlich $1 = P(A_\varepsilon)$ für alle $\varepsilon > 0$.

Fast sichere und L^p-Konvergenz

Für die Folgerung, dass fast sichere Konvergenz L^p-Konvergenz impliziert, benötigen wir mit unseren bisherigen Methoden (majorisierte Konvergenz) eine integrierbare Majorante. Im Folgenden wollen wir diese Forderung abschwächen.

Definition 22. Eine Familie von reellen Zufallsvariablen $(X_i)_{i\in I}$ heißt *gleichgradig integrierbar*, falls

$$\lim_{k\to\infty} \sup_{i\in I} E\left[|X_i| 1_{\{|X_i|>k\}}\right] = 0.$$

Die englische Übersetzung ist uniformly integrable. Anschaulich gesehen heißt diese Bedingung (falls X eine Dichte hat), dass die Dichte am Rand schnell genug abfällt, bzw. dass die Erwartungswerte am Rand schnell genug klein werden. Man beachte, dass gleichgradige Integrierbarkeit stärker ist als Integrierbarkeit.

Beispiel 23. Wir stellen einige wichtige Beispiele vor:

(i) Ist $Y \in L^1$, dann ist Y gleichgradig integrierbar, denn es gilt $|Y|\mathbb{1}_{\{|Y|>k\}} \to 0$ fast sicher und mit majorisierter Konvergenz folgt dann

$$E[|Y|\mathbb{1}_{|Y|>k}] \underset{k\to\infty}{\longrightarrow} 0.$$

(ii) Integrierbare Majorante: Sei $Y \in \mathcal{L}^1$ und $\sup |X_i| < Y$. Dann ist $(X_i)_{i\in I}$ gleichgradig integrierbar, denn nach majorisierter Konvergenz

$$\sup_{i\in I} E\left[|X_i|\mathbb{1}_{\{|X_i|>k\}}\right] \leq E\left[|Y|\mathbb{1}_{\{|Y|>k\}}\right] \underset{k\to\infty}{\longrightarrow} 0.$$

(iii) Jede endliche Familie integrierbarer Zufallsvariablen ist gleichgradig integrierbar, denn $\sup_{1\leq i\leq n} |X_i| \in \mathcal{L}^1$ und

$$\sup_{1\leq i\leq n} E\left[|X_i|\mathbb{1}_{\{|X_i|>k\}}\right] \leq E\Big[\sup_{1\leq i\leq n} |X_i|\mathbb{1}_{\{\sup|X_i|>k\}}\Big] \underset{k\to\infty}{\longrightarrow} 0.$$

(iv) Für $X_n = n\mathbb{1}_{\{U\in[0,\frac{1}{n}]\}}$, $U \sim U[0,1]$ ist für $n > k$

$$E\left[|X_n|\mathbb{1}_{\{|X_n|>k\}}\right] = E[X_n],$$

also ist (X_n) nicht gleichgradig integrierbar.

(v) Gleichgradige Integrierbarkeit ist ein bisschen mehr als Integrierbarkeit, sozusagen eine gleichmäßige $1+\varepsilon$-Integrierbarkeit, wie folgendes Kriterium zeigt: Ist für ein $\varepsilon > 0$

$$\sup_{i\in I} \|X_i\|_{1+\varepsilon} < \infty,$$

so ist

$$\sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|>k\}}] = \sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|^\varepsilon>k^\varepsilon\}}]$$
$$\leq \sup_{i\in I} \frac{E[|X_i|^{1+\varepsilon}]}{k^\varepsilon} \xrightarrow[k\to\infty]{} 0.$$

Lemma 24. *Es sind äquivalent*

(i) $(X_i)_{i\in I}$ *ist gleichgradig integrierbar*

(ii) $\sup_{i\in I} E[|X_i|] < \infty$ *und* $\lim_{\varepsilon\to 0} \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|\mathbb{1}_A] = 0$

(iii) $\lim_{k\to\infty} \sup_{i\in I} E[(|X_i|-k)^+] = 0.$

(iv) *Es gibt* $f:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ *mit* $\frac{f(x)}{x} \xrightarrow[x\to\infty]{} \infty$ *und*

$$\sup_{i\in I} E[f(|X_i|)] < \infty.$$

Beweis: $(i) \Rightarrow (ii)$: Für jedes $\delta > 0$ gibt es ein $k = k(s)$, sodass $\sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|>k\}}] \leq \delta$. Mit $A \in \mathscr{F}$ folgt,

$$E[|X_i|\mathbb{1}_A] = E[|X_i|\mathbb{1}_{A\cap\{|X_i|>k\}}] + E[|X_i|\mathbb{1}_{A\cap\{|X_i|\leq k\}}] \leq \delta + k\cdot P(A).$$

Mit $A = \Omega$ folgt

$$\sup_{i\in I} E[|X_i|] \leq \delta + k < \infty.$$

Außerdem ist

$$\sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|] \leq \delta + k\cdot\varepsilon \xrightarrow[\varepsilon\to 0]{} \delta$$

Da δ beliebig war, folgt (ii).

$(ii) \Rightarrow (iii)$: Zunächst ist $(|X_i|-k)^+ \leq |X_i|\mathbb{1}_{\{|X_i|\geq k\}}$. Sei $\varepsilon > 0$ und $k = k(\varepsilon)$, so dass (mit der Markov-Ungleichung)

$$\sup_{i\in I} P(|X_i| > k) \leq \frac{\sup_{i\in I} E[|X_i|]}{k} < \varepsilon.$$

$$\Rightarrow \qquad \lim_{k\to\infty} \sup_{i\in I} E[(|X_i|-k)^+] = \lim_{\varepsilon\to 0} \sup_{i\in I} E[(|X_i|-k(\varepsilon))^+]$$
$$\leq \lim_{\varepsilon\to 0} \sup_{i\in I} E[|X_i|\mathbb{1}_{\{|X_i|>k(\varepsilon)\}}]$$
$$\leq \lim_{\varepsilon\to 0} \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i|\mathbb{1}_A] = 0.$$

$(iii) \Rightarrow (iv)$: Wir wählen $(k_n) \uparrow \infty$, so dass $\sup_{i\in I} E[(|X_i| - k_n)^+) \leq 2^{-n}$. Setze

$$f(x) = \sum_{n\geq 1} (x - k_n)^+.$$

Dann ist f monoton wachsend und konvex. Dann ist für $x \geq 2k_n$

$$\frac{f(x)}{x} \geq \sum_{k=1}^{n} \big(1 - \frac{k_n}{x}\big) \geq \frac{n}{2},$$

also $\frac{f(x)}{x} \xrightarrow[x\to\infty]{} \infty$. Außerdem gilt wegen monotoner Konvergenz, dass

$$E[f(|X_i|)] = \sum_{n\geq 1} E[(|X_i| - k_n)^+] \leq \sum_{n\geq 1} 2^{-k} = 1.$$

$(iv) \Rightarrow (i)$: Setze $a(k) := \inf_{x\geq k} \frac{f(x)}{x}$, so dass $\lim_{k\to\infty} a(k) = \infty$.

$$\Rightarrow \qquad \sup_{i\in I} E[|X_i| \mathbb{1}_{\{|X_i|\geq k\}}] \leq \frac{1}{a(k)} \sup_{i\in I} E[f(|X_i|) \mathbb{1}_{\{|X_i|\geq k\}}]$$
$$\leq \frac{1}{a(k)} \sup_{i\in I} E[f(|X_i|)] \xrightarrow[k\to\infty]{} 0. \qquad \square$$

Wie handlich die Bedingung (ii) ist, sieht man an folgendem Beispiel:

Beispiel 25. Mit $X \in \mathcal{L}^1$ ist $(X_i)_{i\in J}$ gleichgradig integrierbar genau dann, wenn $(X_i - X)_{i\in I}$ gleichgradig integrierbar ist:

Zunächst ist X gleichgradig integrierbar nach Beispiel 23. Außerdem ist

$$\sup_{i\in I} E[|X_i - X|] \leq E[|X|] + \sup_{i\in I} E[|X_i|] < \infty$$

und

$$\sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i - X| \mathbb{1}_A]$$
$$\leq \sup_{A:P(A)<\varepsilon} \sup_{i\in I} E[|X_i| \mathbb{1}_A] + \sup_{A:P(A)<\varepsilon} E[|X| \mathbb{1}_A] \xrightarrow[\varepsilon\to 0]{} 0$$

und nach Lemma 24 ist $(X_i - X)_{i\in I}$ gleichgradig integrierbar. Die Umkehrung folgt analog.

Das folgende Resultat zeigt, dass aus stochastischer Konvergenz mit der Zusatzbedingung gleichgradiger Integrierbarkeit die L^p-Konvergenz folgt (und umgekehrt).

Konvergenzsatz von Vitali

Theorem 26. *Sei $0 < p < \infty$ und $X_1, X_2, \ldots \in \mathcal{L}^p$. Dann sind äquivalent*

(i) Es gibt $X \in \mathcal{L}^p$ und $X_n \xrightarrow{L^p} X$

(ii) $(|X_n|^p)$ ist gleichgradig integrierbar und es gibt eine Zufallsvariable X mit $X_n \xrightarrow{P} X$.

Gilt (i) oder (ii), so stimmen die Limiten überein.

Beweis. $(i) \Rightarrow (ii)$: Mit der Markov-Ungleichung gilt für alle $\varepsilon > 0$, dass

$$P(|X_n - X| > \varepsilon) \leq \frac{E[|X_n - X|^p]}{\varepsilon^p} = \frac{\|X_n - X\|_p^p}{\varepsilon^p} \longrightarrow 0,$$

so dass $X_n \xrightarrow{P} X$.

Nun verwenden wir Lemma 24: Sei $\varepsilon > 0$ und $N = N(\varepsilon)$ so, dass $\|X_n - X\|_p < \varepsilon \; \forall n \geq N$. Dann folgt

$$\begin{aligned} \sup E[|X_n|^p]^{1\wedge\frac{1}{p}} &= \sup \|X_n\|_p^{p\wedge 1} \\ &\leq \sup_{n<N} \|X_n\|_p^{p\wedge 1} + \underbrace{\sup_{n\geq N} \|X_n - X\|_p^{p\wedge 1}}_{\leq \varepsilon^{p\wedge 1}} + \|X\|_p^{p\wedge 1} < \infty \end{aligned}$$

unter Verwendung der Minkowski-Ungleichung.

Für $A \in \mathscr{F}$ erhalten wir ebenso

$$\begin{aligned} \sup \big(E[|X_n|^p \mathbb{1}_A]\big)^{1\wedge\frac{1}{p}} &= \sup \|X_n \mathbb{1}_A\|_p^{p\wedge 1} \\ &\leq \sup_{n<N} \|X_n \mathbb{1}_A\|_p^{p\wedge 1} + \underbrace{\sup_{n\geq N} \|(X_n - X)\mathbb{1}_A\|_p^{p\wedge 1}}_{\leq \varepsilon^{p\wedge 1}} + \|X\mathbb{1}_A\|_p^{p\wedge 1}. \end{aligned}$$

Da $\|X\mathbb{1}_A\|_p \to 0$ für $P(A) \to 0$, folgt

$$\lim_{\delta\to 0} \sup_{A:P(A)<\delta} \sup_{n\in\mathbb{N}} E(|X_n|^p \mathbb{1}_A) \leq \varepsilon^{2p\wedge 1}$$

Die Behauptung folgt, weil $\varepsilon > 0$ beliebig war.

$(ii) \Rightarrow (i)$ Aus $X_n \xrightarrow{P} X$ folgt die Existenz einer Teilfolge (n_k) so dass $X_{n_k} \xrightarrow{\text{f.s.}} X$. Außerdem ist $\sup E[|X_n|^p] < \infty$ nach Lemma 24. Nach dem Lemma von Fatou ist

$$E[|X|^p] = E\big[\liminf_{k\to\infty} |X_{n_k}|^p\big] \leq \sup_{n\in\mathbb{N}} E[|X_n|^p] < \infty$$

und somit $X \in \mathcal{L}^p$. Weiterhin ist $|X_n - X|^p$ gleichgradig integrierbar (siehe Beispiel 25).

Wieder wenden wir Lemma 24 an. Geeignete A finden sich, da

$$P(|X_n - X| > \delta) \underset{n\to\infty}{\longrightarrow} 0$$

und somit

$$\begin{aligned} \lim_{n\to\infty} E[|X_n - X|^p] &= \lim_{n\to\infty} E[|X_n - X|^p \mathbb{1}_{\{|X_n - X| > \delta\}}] \\ &\quad + \lim_{n\to\infty} \underbrace{E[|X_n - X|^p \mathbb{1}_{\{|X_n - X| \leq \delta\}}]}_{\leq \delta^p} \\ &\leq \delta^p \end{aligned}$$

und die Behauptung folgt, weil δ beliebig war. □

Korollar 27. *Seien $0 < p < \infty$, $X_1, X_2, \dots \in \mathcal{L}^p$ und $X_n \overset{P}{\to} X$. Dann sind äquivalent*

(i) $X_n \overset{L^p}{\to} X$
(ii) $\|X_n\|_p \to \|X\|_p$
(iii) $(|X_n|^p)_{n\geq 1}$ *ist gleichgradig integrierbar*

Dieses nützliche Kriterium werden wir immer wieder verwenden. Äquivalent zu (ii) ist natürlich

$$E[|X_n|^p] \to E[|X|^p]$$

Beweis. Die Äquivalenz von (i) und (iii) haben wir ja gerade gezeigt. Es bleiben: (i) $\Rightarrow$ (ii): Mit der (umgekehrten) Minkowski-Ungleichung erhalten wir

$$0 \leq \big| \| X_n \|_p - \| X \|_p \big| \leq \| X_n - X \|_p \to 0,$$

wegen L^p-Konvergenz.

Für (ii)$\Rightarrow$(iii): ÜA (Blatt 5 Aufgabe 1) □

Das starke Gesetz der großen Zahl

Ein zentrales Konzept dieses Abschnittes ist die Unabhängigkeit von Zufallsvariablen. Wir werden in diesem Zusammenhang auch die wichtigen Hilfsmittel von Borel-Cantelli und das Kolmogorov'sche 0-1-Gesetz kennenlernen.

Unabhängigkeit und 0-1 Gesetze

Die zentrale Eigenschaft ist *Unabhängigkeit*. Ausgehend von der Unabhängigkeit für beliebige Familien von Mengen, welche man zurückführt auf jeweils endliche Kombinationen von Mengen, führt man die Unabhängigkeit von Mengensystem analog ein und definiert die Unabhängigkeit von Zufallsvariablen schließlich über die Unabhängigkeit der von ihnen erzeugten σ-Algebren.

Definition 28. (i) Eine Familie messbarer Mengen $(A_i)_{i\in I}$ heißt *unabhängig*, falls

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j) \tag{29}$$

für alle $J \subset I$ endlich.

(ii) Eine Familie von Mengensystemen $(C_i)_{i\in I}$, $C_i \subseteq \mathscr{F}\ \forall i \in I$ heißt unabhängig, falls (29) für alle $J \subseteq I$, J endlich und $A_j \in C_j$, $j \in J$, gilt.

(iii) Eine Familie von Zufallsvarialben $(X_i)_{i\in I}$ heißt unabhängig, falls $(\sigma(X_i))_{i\in I}$ unabhängig sind.

Man erhält unmittelbar, dass $(X_i)_{i\in I}$ unabhängig sind, falls sich ihre (endlichen) gemeinsamen Verteilungen als Produkt von den Marginalien ergibt, d.h.

$$P(X_i \in A_i, i \in J) = \prod_{i\in J} P(X_i \in A_i)$$

für alle $J \subset I, A_i \in \mathscr{F}$, J endlich.

Als wichtiges Hilfsmittel erweist sich das folgende „Blockungslemma“:

Lemma 30. *Seien $(\Omega_i', \mathscr{F}_i')$, $(\Omega_i'', \mathscr{F}_i'')$, $i \in J$, Meßräume und $(X_i)_{i\in J}$ unabhängige Zufallsvariablen mit $X_i : \Omega \to \Omega_i'$. Sind $\varphi_i : \Omega_i' \to \Omega_i''$ messbar, $i \in J$, so ist*

$$(\varphi_i(X_i))_{i\in J} \qquad \text{unabhängig.}$$

Beweis. Zunächst ist $\sigma(\varphi_i(X_i)) \subset \sigma(X_i)$, da φ_i messbar. Da $\sigma(X_i)_{i\in J}$ bereits unabhängig sind, folgt die Behauptung. □

Zwei Zufallsvariablen heißen **unkorreliert**, falls $\mathrm{Cov}(X, Y) = 0$. Wir erhalten leicht, dass Unabhängigkeit Unkorreliertheit impliziert.

Satz 31. *Sind $X, Y \in \mathcal{L}^1$ unabhängig, so ist $X \cdot Y \in \mathcal{L}^1$ und*

$$E[XY] = E[X] \cdot E[Y].$$

Beweis. Die Aussage gilt für $X = \mathbb{1}_A$ und $Y = \mathbb{1}_B$. Mit Linearität folgt die Aussage für einfache X, Y mit monotoner Konvergenz für Positiv-/Negativteile und somit die Behauptung. □

Beispiel 32. Ist $X \sim \mathcal{N}(0, 1)$, so sind X und X^2 unkorreliert aber nicht unabhängig.

Sind $A_1, A_2, \dots$ Ereignisse, so definiert der folgende Limes Superior das Ereignis „unendlich viele der (A_n) treten ein“. Für $A_1, A_2, \dots \in J$ definieren wir

$$\begin{aligned} \limsup_{n\to\infty} A_n &:= \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m \\ &= \{\omega \in \Omega | \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\}. \end{aligned}$$

Theorem 33 (Borel-Cantelli). *(i) Seien $A_1, A_2, \dots \in \mathscr{F}$, so gilt*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 0.$$

(ii) Sind die $(A_n)_{n\geq 1}$ unabhängig, so gilt

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 1.$$

Beweis. Wir zeigen zunächst den ersten Teil: Stetigkeit von oben von P impliziert[17]

$$P\Big(\limsup_{n\to\infty} A_n\Big) = \lim_{n\to\infty} P\Big(\bigcup_{m\geq n} A_n\Big) \leq \lim_{n\to\infty} \sum_{m\geq n} P(A_m) = 0.$$

[17] Dies erhalten wir aus monotoner Konvergenz, oder elementarer: Aus der σ-Additivität von Maßen, siehe Satz A.14 und Gleichung (5).

Nun kommen wir zum zweiten Teil. Für $x \in [0,1]$ gilt: $\log(1-x) \leq -x$ also folgt mit Stetigkeit von unten, dass

$$\begin{aligned} P((\limsup A_n)^c) &= P\Big(\bigcup_{n\geq 1}\bigcap_{m\geq n} A_m^c\Big) \\ &= \lim_{n\to\infty} \exp\Big(\sum_{m\geq n} \log\left(1-P(A_m)\right)\Big) \\ &\leq \lim_{n\to\infty} \exp\Big(-\sum_{m\geq n} P(A_m)\Big) = 0 \end{aligned}$$

und der Beweis ist fertig. □

Beispiel 34 (Unendlicher Münzwurf)**.** $X_1, X_2, \ldots$ iid, $X_i \in \{0,1\}$ und $P(X_n = 1) = \dfrac{1}{2} \Rightarrow \sum\limits_{n=1}^{\infty} P(A_n) = \infty \Rightarrow$ fast sicher erscheint unendlich mal Kopf.

Auf die Unabhängigkeit kann nicht verzichtet werden: Ist

$$B_1 = B_2 = B_n = \{X_1 = \text{ Kopf}\},$$

so gilt $\sum P(B_n) = \infty$, aber $P\big(\limsup\limits_{n\to\infty} B_n\big) = P(X_1 = \text{ Kopf}) = \dfrac{1}{2}$.

Terminale Ereignisse

Aussagen wie „unendlich oft tritt Kopf (bzw. ein Ereignis) ein“ bilden eine spezielle Klasse von Ereignissen, die terminalen Ereignisse, die wir nun genauer untersuchen.

Zunächst einmal ein Satz der von unabhängingen Mengensystem auf die Unabhängigkeit der davon erzeugten σ-Algebren schließen lässt. Hierfür braucht es aber eine Zusatzbedingung: Die Mengensystem müssen *durchschnittsstabil* sein.

Satz 35. *Sei $(C_i)_{i\in I}$ eine Familie unabhängiger Mengensysteme (messbarer Mengen). Ist jedes C_i, $i \in I$, durchschnittsstabil, so sind auch $(\sigma(C_i))_{i\in I}$ unabhängig.*

Beweis. Sei $J = \{j_1, \ldots, j_n\} \subset I$ und ohne Einschränkung $|J| > 1$. Wir wählen $A_{j_i} \in C_{j_i}$, $i = 2, \ldots, n$ und definieren

Einmal wieder verwenden wir den Dynkin-Trick: Wir suchen uns gute Mengen und zeigen dass es sich um ein Dynkin-System handelt.

$$\mathscr{D} := \Big\{A \in \mathscr{F} : P(A \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A) \cdot \prod_{i=2}^{n} P(A_{j_i})\Big\}.$$

Dann ist $\mathscr{D}$ Dynkin-System, denn

1) $A \subseteq B, A, B \in \mathscr{D} \Rightarrow B \setminus A \in \mathscr{D}$, da

$$\begin{aligned} P\Big(B \setminus A \cap \underbrace{\bigcap_{i=2}^n A_{j_i}}_{:=C}\Big) &= P(B \cap C) - P(A \cap C) \\ &= (P(B) - P(A)) \prod_{i=2}^n P(A_{j_i}) \\ &= P(B \setminus A) \cdot \prod_{i=2}^n P(A_{j_i}). \end{aligned}$$

2) Für $A_1, A_2, \ldots \in \mathscr{D}$ mit $A_1 \subseteq A_2 \subseteq \cdots$ ist

σ-Stetigkeit von unten

$$\begin{aligned} P\Big(\bigcup_{n=1}^\infty A_n \cap C\Big) &= \lim_{m\to\infty} P(A_m \cap C) \\ &= \lim_{n\to\infty} P(A_m) \cdot \prod_{i=2}^n P(A_{j_i}) \\ &= P\Big(\bigcup_{n=1}^\infty A_n\Big) \cdot \prod_{i=2}^n P(A_{j_i}). \end{aligned}$$

Da C_{j_1} durchschnittstabil ist, ist $\sigma(C_{j_1}) \subseteq \mathscr{D}$, also folgt

$$P(A \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A) \cdot \prod_{i=2}^n P(A_{j_i})$$

für alle $A \in \sigma(C_{j_1})$. Da die Wahl von j_1 beliebig war, erhält man die Aussage. $\square$

Bemerkung 36. (i) Als leichte Folgerung erhält man, dass $(A_i)_{i\in I}$ genau dann unabhängig sind, falls $(\mathbb{1}_{A_i})_{i\in I}$ unabhängig sind.

(ii) Ist $(\mathscr{F}_i)_{i\in I}$ eine Familie unabhängiger σ-Algebren und $\mathcal{I} = \{I_k : k \in K\}$ eine Partition von I, so sind auch $(\sigma(\mathscr{F}_i : i \in I_k) : k \in K)$ unabhängig.

Definition 37. Sei $\mathbb{F} := (\mathscr{F}_1, \mathscr{F}_2, \ldots)$ eine Folge von σ-Algebren. Dann heißt

$$\mathcal{T}(\mathbb{F}) := \bigcap_{n\geq 1} \sigma\Big(\bigcup_{m\geq n} \mathscr{F}_n\Big)$$

die σ-Algebra der *terminalen Ereignisse* von $\mathbb{F}$. Eine σ-Algebra $\mathscr{G}$ heißt *trivial*, falls $P(A) \in \{0,1\}$ für alle $G \in \mathscr{G}$.

Lemma 38. *(i) $\mathscr{G}$ ist trivial $\Leftrightarrow$ $\mathscr{G}$ ist zu sich selbst unabhängig.*

(ii) Sei $\mathscr{G}$ trivial, $\mathcal{X}$ ein separabler metrischer Raum, X Zufallsvariable mit Werten in $\mathcal{X}$. Dann gilt: Ist X $\mathscr{G}$-messbar, so ist X P-fast sicher konstant.

Beweis. (i) Seien $A, B \in \mathscr{G}$. Dann ist

$$\begin{aligned} P(A \cap B) &= \begin{cases} 0 = P(A) \cdot P(B), & \text{falls } P(A) = 0 \text{ oder } P(B) = 0 \\ 1, & \text{falls } P(A) = P(B) = 1 \end{cases} \\ &= P(A) \cdot P(B). \end{aligned}$$

Also ist $\mathscr{G}$ von sich selbst unabhängig. Umgekehrt folgt aus $\mathscr{G}$ ist zu sich selbst unabhängig, dass $P(A) = P(A \cap A) = P(A)^2 \in \{0,1\}$.

(ii) Für $n \in \mathbb{N}$ seien $(A_i^n)_{i=1,2,\dots}$ eine abzählbare Überdeckung von E mit Bällen vom Radius $\frac{1}{n}$. Da $\mathcal{G}$ trivial ist, folgt

$$P(X \in A_i^n) \in \{0,1\}.$$

Wir setzen $I_n = \{i \in \mathbb{N} : P(X \in A_i^n) = 1\} \neq \varnothing$.

$$\overset{\text{Stetigkeit von oben}}{\Longrightarrow} \qquad P\Big(X \in \bigcap_{n=1}^{\infty} \bigcap_{i \in I_n} A_i^n\Big) = 1.$$

Wobei wir verwendet haben, dass aus $P(A) = 1$ und $P(B) = 1$ folgt, dass $P(A \cap B) = 1$. Allerdings hat $\bigcap_{n=1}^{\infty} \bigcap_{i \in I_n} A_i^n$ höchsten ein Element und die Behauptung folgt. □

Die Hauptaussage dieses Abschnitts ist das Kolmogorov'sche 0-1-Gesetz. Das Borel-Cantelli Lemma ist ebenfalls ein 0-1-Gesetz. Ein ähnliches Gesetz ist das 0-1-Gesetz von Hewitt Savage, dass man aus dem Kolmogorvschen herleiten kann.

Theorem 39. *Sei $\mathbb{F} := (\mathscr{F}_1, \mathscr{F}_2, \dots)$ eine Folge* unabhängiger *σ-Algebren. Dann ist die terminale σ-Algebra $\mathcal{T}(\mathbb{F})$ trivial.*

0-1- Gesetz von Kolmogorov

Der Beweis beruht auf einer besonders schönen Idee, braucht aber einige Vorarbeit - die wir bereits erbracht haben ! Nun geht der Beweis erstaunlich schnell.

Beweis. Sei $\mathcal{T}_n := \sigma\Big(\bigcup_{m>n} \mathscr{F}_m\Big)$, $n = 0, 1, 2, \dots$.

Nach Bemerkung 36(ii) sind auch $(\mathscr{F}_1, \dots, \mathscr{F}_n, \mathcal{T}_n)$ unabhängig, $n = 1, 2, \dots$ und somit auch $(\mathscr{F}_1, \dots, \mathscr{F}_n, \mathcal{T})$, $n = 1, 2, \dots$.

Da dies für alle n gilt, sind auch $(\mathcal{T}, \mathscr{F}_1, \mathscr{F}_2, \dots)$ unabhängig.

Wieder mit Bemerkung 36(ii) folgt, dass $(\mathcal{T}_0, \mathcal{T})$ unabhängig sind.

Da nun $\mathcal{T} \subseteq \mathcal{T}_0$ ist folgt, dass $\mathcal{T}$ von sich selbst unabhängig und somit trivial. $\square$

Bemerkung 40. Ein typisches Anwendungsbeispiel des Komogorov'schen 0-1 Gesetzes ist die Aussage, dass fast sichere Grenzwerte von unabhängigen Zufallsvariablen konstant sein müssen. Hierzu zeigt man zunächst dass $\{\lim X_n \text{ konvergiert}\}$ ein terminales Ereignis ist. Als nächstes zeigt man, dass der Limes messbar bezüglich der terminalen σ-Algebra ist und somit folgt, dass er trivial sein muss.

Satz 41. *Seien $X_1, X_2, \dots$ unabhängig, $S_n = X_1 + \dots + X_n$. Dann gilt*

$$P\big(S_n \textit{ konvergiert für } n \to \infty\big) \in \{0, 1\}$$
$$P\big(n^{-1} S_n \textit{ konvergiert für } n \to \infty\big) \in \{0, 1\}$$

Beweis. Mit (X_i) sind auch $\mathbb{F} = (\mathscr{F}_i)$ mit $\mathscr{F}_i = \sigma(X_i)$, $i = 1, 2, \dots$ unabhängig. Die zentrale Beobachtung ist, dass die Menge

$$\{\omega : S_n(\omega) \text{ konvergiert für } n \to \infty\}$$

messbar ist bezüglich der terminalen σ-Algebra $\mathcal{T}(\mathbb{F})$: In der Tat, S_n konvergiert genau dann wenn $S_n - X_1 - \dots - X_m$ konvergiert. Damit folgt die erste Aussage mit dem 0-1-Gesetzt von Komogorov.

Ebenso sieht man, dass $n^{-1} S_n$ genau dann konververgiert wenn $n^{-1}(S_n - X_1 - \dots - X_m)$ konvergiert. $\square$

Die Maximalungleichung

Eine wichtige und starke Ungleichung ist die folgende Maximal-Ungleichung von Kolmogorov. Sie wird uns auch später noch in allgemeineren Varianten begegnen. Sie stärkt das schwache Gesetz der großen Zahl auf eine gleichmäßige Aussage.

Die Maximal-Ungleichung von Kolmogorov.

Satz 42. *Seien $X_1, X_2, \dots \in L^2$ unabhängig. Dann gilt für jedes $\varepsilon > 0$, dass*

$$P\Big(\sup_{n \in \mathbb{N}} \Big| \sum_{k=1}^n X_k - E[X_k] \Big| > \varepsilon\Big) \leq \frac{\sum_{n=1}^\infty \operatorname{Var}(X_n)}{\varepsilon^2}.$$

Beweis. Ohne Einschränkung können wir annehmen, dass $E[X_k] = 0$ für alle k. Wie zuvor betrachten wir $S_n = X_1 + \cdots + X_n$ und setzen

$$T := \inf\{n : |S_n| > \varepsilon\}.$$

Die zentrale Beobachtung ist

$$P(\sup_n |S_n| > \varepsilon) = P(T < \infty).$$

Das Blockungslemma garantiert, dass für $k \leq n$ die beiden Zufallsvariablen $S_k \cdot \mathbb{1}_{\{\}} T = k$ und $S_n - S_k$ unabhängig sind. Es folgt, dass

$$\begin{aligned}
\sum_{k=1}^{n} \operatorname{Var}(X_k^2) = E[S_n^2] &\geq \sum_{k=1}^{n} E[S_n^2 \mathbb{1}_{\{T=k\}}] \\
&= \sum_{k=1}^{n} E[(S_n - S_k + S_k)^2 \mathbb{1}_{\{T=k\}}] \\
&= \sum_{k=1}^{n} E[S_k^2 \mathbb{1}_{\{T=k\}}] + 2E[S_k (S_n - S_k)\mathbb{1}_{\{T=k\}}] + E[(S_n - S_k)^2 \mathbb{1}_{\{T=k\}}] \\
&= \sum_{k=1}^{n} E[S_k^2 \mathbb{1}_{\{T=k\}}] + E[(S_n - S_k)^2 \mathbb{1}_{\{T=k\}}] \\
&\geq \sum_{k=1}^{n} E[S_k^2 \mathbb{1}_{\{T=k\}}] \geq \varepsilon^2 P(T \leq n)
\end{aligned}$$

Wobei wir für die letzte Ungleichung die Definition von T verwendet haben. Nun folgt die Behauptung mit $n \to \infty$. □

Kombinieren wir die Maximal-Ungleichung mit unseren Konvergenzkriterien für fast-sichere Konvergenz erhalten wir folgenden Konvergenzsatz. Bemerkenswert ist hieran, dass nicht i.i.d. vorausgesetzt wird, sondern lediglich Unabhängigkeit (und eben Summierbarkeit der Varianzen).

Lemma 43. *Seien $X_1, X_2, \cdots \in L^2$ unabhängig und es gelte $\sum_{n=1}^{\infty} \operatorname{Var}(X_n) < \infty$. Dann konvergiert $\sum_{k=1}^{n} (X_k - E[X_k])$ fast sicher.*

Ein nützliches Konvergenzkriterium für die fast sichere Konvergenz von Reihen zentrierter Zufallsvariablen.

Beweis. Wieder sei ohne Einschränkung $E[X_k] = 0, k = 1, 2, \ldots$ und $S_n = X_1 + \cdots + X_n$. Für $\varepsilon > 0$ gilt nach Satz 42

$$\lim_{k\to\infty} P(\sup_{n\geq k} |S_n - S_k| > \varepsilon) \leq \lim_{k\to\infty} \frac{\sum_{n=k+1}^{\infty} E[X_n^2]}{\varepsilon^2} = 0.$$

Und somit konvergiert $\sup_{n\geq k} |S_n - S_k| \xrightarrow{P} 0$. Damit gibt es eine Teilfolge $k_1, k_2, \ldots$ mit $\sup_{n\geq k_i} |S_n - S_{k_i}| \xrightarrow{f.s.} 0$.

Da aber $(\sup_{n\geq k} |S_n - S_k|)_{k=1,2,\ldots}$ fallend ist, gilt $\sup_{n\geq k} |S_n - S_k| \xrightarrow{f.s.} 0$, was gleichbedeutend damit ist, das (S_n) konvergiert. □

Bemerkung 44. Im obigen Beweis haben wir folgende Aussagen verwendet:

1. Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge mit $a_n \geq 0$ und $\sum_{n=1}^\infty a_n < \infty$, dann folgt $\lim_{k\to\infty}\sum_{n=k}^\infty a_n = 0$.
2. Sei $(a_n)_{n\in\mathbb{N}}$ eine monotone Folge und es gebe eine Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$, sodass $a_{n_k} \to a$, dann folgt $a_n \to a$.

Der Beweis dieser Aussagen kann als Übungsaufgabe geführt werden.

Oft wollen wir eine Folge von Zufallsvariablen nur ein bisschen verändern, ohne dass sich das Konvergenzverhalten ändert.

Definition 45. Zwei Folgen von Zufallsvariablen (X_n) und (Y_n) heißen *asymptotisch äquivalent*, falls

$$\sum_{n\geq 1} P(X_n \neq Y_n) < \infty. \tag{46}$$

Bemerkung 47. Mit dem Satz von Borell-Cantelli folgt sofort, dass aus f.s. Konvergenz von (Y_n) diejenige von (X_n) folgt falls sie asymptotisch äquivalent sind: Denn (46) impliziert, dass

$$P(\limsup\{X_n \neq Y_n\}) = 0.$$

D.h. es existiert eine Menge A mit Wahrscheinlichkeit eins und eine Zufallsvariable $n_0 \in \mathbb{N}$, so dass für alle $\omega \in A$ und $n \geq n_0(\omega)$ gilt, dass $X_n(\omega) = Y_n(\omega)$. Dann folgt, dass $\sum_{n\geq 1}(X_n - Y_n)$ fast sicher konvergiert.

Der Drei-Reihensatz

Zentrales Konvergenzkriterium wird der folgende *Drei-Reihensatz* sein:

Satz 48. *Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen. Dann konvergiert $S_n = \sum_{k=1}^n X_k$ genau dann P-fast sicher, falls ein $c \in (0,\infty)$ existiert, so dass die folgenden drei Reihen konvergieren:*

$$(i)\ \sum_{n\geq 1} P(|X_n| < c), \qquad (ii)\ \sum_{n\geq 1} \operatorname{Var}(X_n \mathbb{1}_{\{|X_n|\geq c\}}) \qquad (iii)\ \sum_{n\geq 1} E\big[X_n \mathbb{1}_{\{|X_n|\geq c\}}\big].$$

Beweis. Betrachten wir $Y_n = X_n \mathbb{1}_{\{|X_n| \geq c\}} - E[X_n \mathbb{1}_{\{|X_n| \geq c\}}]$. Dann impliziert Bedingung (ii) mit Lemma 43, dass $\sum_{k=1}^n Y_k$ fast sicher konvergiert. Mit (iii) erhalten wir, dass auch

$$\sum_{k=1}^{n} X_k \mathbb{1}_{\{|X_k| \geq c\}} = \sum_{k=1}^{n} Y_k + \sum_{k=1}^{n} E[X_k \mathbb{1}_{\{|X_k| \geq c\}}]$$

konvergiert.

Bedingung (i) impliziert, dass (X_n) und (Y_n) asymptotisch äquivalent sind, denn

$$\sum_{n \geq 1} P\big(X_n \neq X_n \mathbb{1}_{\{|X_n| \geq c\}}\big) = \sum_{n \geq 1} P(|X_n| < c).$$

Nach obiger Bemerkung konvergiert also auch $\sum_{k=1}^n X_n$. □

Das starke Gesetz der großen Zahl

Wir erhalten das starke Gesetz der großen Zahl von Kolmogorov unter der Annahme über existierende zweite Momente. Im Folgenden soll diese Annahme noch so weit wie möglich abgeschwächt werden.

Wir beginnen mit zwei nützlichen Hilfsmitteln aus der Analysis. Für einen Beweis: Siehe Henze[18], Kapitel 6.2.

[18] Norbert Henze. *Stochastik für Einsteiger*. Springer, 2013

(i) **Lemma von Kronecker:** Seien $x_1, x_2, \dots$ reelle Zahlen, $a_1, a_2, \dots$ positive Zahlen mit $a_n \uparrow \infty$ so, dass $\sum \frac{x_n}{a_n}$ konvergiert. Dann gilt, dass
$$\frac{1}{a_n} \sum_{k=1}^{n} x_k \longrightarrow 0.$$

(ii) **Lemma von Cesàro:** Sind $b_1, b_2, \dots$ reelle Zahlen mit $b_n \to b \in \mathbb{R}$ und $a_1, a_2, \dots$ monoton wachsende, positive reelle Zahlen mit $a_n \uparrow \infty$ und $a_0 = b_0 = 0$, so gilt, dass
$$lim_{n\to\infty} \frac{1}{a_n} \sum_{i=1} (a_j - a_{j-1}) b_{j-1} = b.$$

Bemerkung 49 (Cesaro)**.** Das Lemma von Cesàro nutzen wir meistens in der folgenden Weise: konvergieren $x_1, x_2, \dots$ gegen ein x, so konvergiert auch das arithmetische Mittel gegen x.

$$x_n \to x \;\Rightarrow\; \frac{1}{n} \sum_{i=1}^{n} x_i \to x$$

Theorem 50 (Starkes Gesetz großer Zahlen)**.** *Seien $X_1, X_2, \dots \in L^2$ unabhängige und reellwertige Zufallsvariablen. Gibt es eine Folge positiver Zahlen $a_n \uparrow \infty$, so dass*

$$\sum_{n \geq 1} \frac{\mathrm{Var}(X_n)}{a_n^2} < \infty,$$

so folgt, dass

$$\frac{1}{a_n}\sum_{k=1}^{n}(X_k - E[X_k]) \longrightarrow 0$$

P-fast sicher.

Beweis. Wir setzen $Y_n = (a_n)^{-1}(X_n - E[X_n])$. Dann sind $Y_1, Y_2, \ldots$ unabhängig, zentriert und $\sum \operatorname{Var}(Y_n) < \infty$ nach Voraussetzung. Nach Lemma 43 konvergiert $\sum_{k=1}^{n} Y_k$ P-fast sicher.

Mit dem Kronecker Lemma folgt, dass

$$\frac{1}{a_n}\sum_{k=1}^{n} a_k Y_k = \frac{1}{a_n}\sum_{k=1}^{n}(X_k - E[X_k]) \longrightarrow 0$$

P-fast sicher. □

Nun beweisen wir das SGGZ unter minimaler L^1-Voraussetzung.

Theorem 51. *Seien $X_1, X_2, \ldots$ reellwertige Zufallsvariablen und i.i.d. Gilt $E[|X_1|] < \infty$, so folgt, dass*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow E[X_1],$$

P-fast sicher.

Das starke Gesetz der großen Zahl unter minimaler Integrabilitäts-Voraussetzung. Es ist allerdings auch klar, dass man die Annahme der identischen Verteilung noch weiter abschwächen kann.

Für den Fall, dass $E[|X_1|] = \infty$ folgt Divergenz, siehe etwa [19], Satz 2.6.13.

[19] Ludger Rüschendorf. *Wahrscheinlichkeitstheorie*. Springer, 2016

Beweis. Die Idee für den Beweis ist, die Zufallsvariablen geschickt abzuschneiden. Hierzu setzen wir $Y_n = X_n \mathbb{1}_{\{|X_n| \leq n\}}$, $n \geq 1$. Wir zeigen, dass (X_n) und (Y_n) asymptotisch äquivalent sind, denn

$$\begin{aligned}
\sum_{n\geq 1} P(X_n \neq Y_n) &= \sum_{n\geq 1} P(|X_n| > n) = \sum_{n\geq 1} P(|X_1| > n) \\
&= \sum_{n\geq 1}\sum_{j\geq n} P(j < |X_1| \leq j+1) \\
&= \sum_{j\geq 1}\sum_{n=1}^{j} P(j < |X_1| \leq j+1) = \sum_{j\geq 1} jP(j < |X_1| \leq j+1) \\
&\leq E[|X_1|] < \infty
\end{aligned}$$

Nun kann man Theorem 50 anwenden, und zwar auf (Y_n). In der Tat, es gilt

$$\begin{aligned}
\sum_{n\geq 1} \frac{Var(Y_n)}{n^2} \leq \sum_{n\geq 1} \frac{E[Y_n^2]}{n^2} &= \sum_{n\geq 1} \frac{1}{n^2}\int_{\{|X_1|\leq n\}} (X_1)^2 dP \\
&= \sum_{n\geq 1} \frac{1}{n^2}\sum_{j=1}^{n}\int_{\{j-1<|X_1|\leq j\}} (X_1)^2 dP \\
&= \sum_{j\geq 1}\int_{\{j-1<|X_1|\leq j\}} (X_1)^2 dP \cdot \sum_{n=j}^{\infty}\frac{1}{n^2} \\
&\leq \sum_{j\geq 1}\int_{\{j-1<|X_1|\leq j\}} j\cdot|X_1| dP \cdot \sum_{n=j}^{\infty}\frac{1}{n^2}.
\end{aligned}$$

Die letzte Reihe können wir direkt auswerten: Zunächst ist für $j = 1$, $\sum_{n\geq 1}\frac{1}{n^2} = \frac{\pi^2}{6}$. Für $j > 1$ ist

$$\sum_{n=j}^{\infty}\frac{1}{n^2} \leq \int_{(j-1,\infty)} \frac{1}{x^2}dx = -\frac{1}{x}\Big|_{j-1}^{\infty} = \frac{1}{j-1}.$$

Multiplizieren wir diese Reihe mit j, so erhalten wir zwei als obere Schranke:

$$j \sum_{n=j}^{\infty} \frac{1}{n^2} \leq \begin{cases} \frac{\pi^2}{6} \leq 2 & j = 1 \\ \frac{j}{j-1} \leq 2 & j > 1. \end{cases}$$

Es folgt, dass

$$\sum_{n \geq 1} \frac{Var(Y_n)}{n^2} \leq 2E[|X_1|] < \infty.$$

Theorem 50 impliziert nun, dass

$$\frac{1}{n} \sum_{j=1}^{n} (Y_k - E[Y_k]) \longrightarrow 0$$

P-fast sicher, wobei $E[Y_k] \to E[X_1]$ mit Hilfe von majorisierter Konvergenz. Nutzt man Bemerkung 47 zusammen mit Bemerkung 49, so folgt die Behauptung. □

Das Glivenko-Cantelli Lemma

In der nichtparametrischen Statistik spielt die empirische Verteilungsfunktion eine zentrale Rolle. Sie ist ein natürlicher Kandidat für die Schätzung der Verteilung von austauschbaren Zufallsvariablen - insbesondere für unabhängige und identisch verteilte Zufallsvariabelen. Man beachte, dass jeder Zufallsvariable das gleiche Gewicht zugewiesen wird.

Definition 52. Seien $X_1, X_2, \ldots$ Zufallsvariable. Dann heißt

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

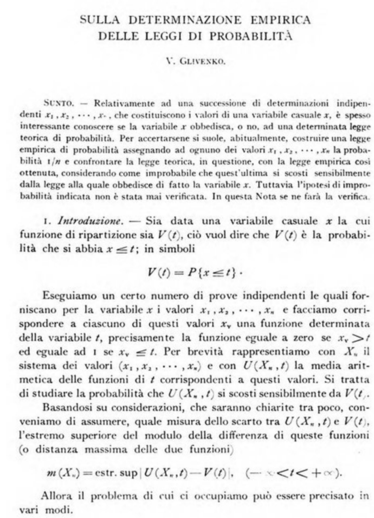
die *empirische Verteilung* von $X_1, \ldots, X_n$. Sind die Zufallsvarialbe reellwertig, so heißt

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\{X_i \leq x\}}$$

die *empirische Verteilungsfunktion*.

Das folgende Resultat ist ein überraschend starkes Resultat - es zeigt, dass die empirische Verteilungsfunktion gleichmäßig gegen die wahre Verteilungsfunktion von iid Zufallsvariablen konvergiert und zwar auf einer Menge mit Wahrscheinlichkeit 1. Dieses zentrale Resultat bildet den Grundstein der gesamten nichtparametrischen Statistik.

Das folgende Theorem wird auch als Hauptsatz der Statistik bezeichnet. Es wurde bereits 1933 von Glivenko und Cantelli bewiesen, beide Arbeiten sind im Giornale dell'Istituto italiano degli attuari erschienen und im Internet abrufbar (in Italienisch).



SULLA DETERMINAZIONE EMPIRICA
DELLE LEGGI DI PROBABILITÀ

V. GLIVENKO.

SUNTO. — Relativamente ad una successione di determinazioni indipendenti $x_1, x_2, \cdots, x_n$, che costituiscono i valori di una variabile casuale x, è spesso interessante conoscere se la variabile x obbedisca, o no, ad una determinata legge teorica di probabilità. Per accertarsene si suole, abitualmente, costruire una legge empirica di probabilità assegnando ad ognuno dei valori $x_1, x_2, \cdots, x_n$ la probabilità $1/n$ e confrontare la legge teorica, in questione, con la legge empirica così ottenuta, considerando come improbabile che quest'ultima si scosti sensibilmente dalla legge alla quale obbedisce di fatto la variabile x. Tuttavia l'ipotesi di improbabilità indicata non è stata mai verificata. In questa Nota se ne farà la verifica.

1. *Introduzione.* — Sia data una variabile casuale x la cui funzione di ripartizione sia $V(t)$, ciò vuol dire che $V(t)$ è la probabilità che si abbia $x \leq t$; in simboli

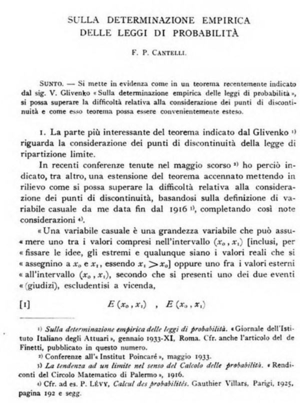
$$V(t) = P\{x \leq t\} \cdot$$

Eseguiamo un certo numero di prove indipendenti le quali forniscano per la variabile x i valori $x_1, x_2, \cdots, x_n$ e facciamo corrispondere a ciascuno di questi valori x_ν una funzione determinata della variabile t, precisamente la funzione eguale a zero se $x_\nu > t$ ed eguale ad 1 se $x_\nu \leq t$. Per brevità rappresentiamo con X_n il sistema dei valori $(x_1, x_2, \cdots, x_n)$ e con $U(X_n, t)$ la media aritmetica delle funzioni di t corrispondenti a questi valori. Si tratta di studiare la probabilità che $U(X_n, t)$ si scosti sensibilmente da $V(t)$.

Basandosi su considerazioni, che saranno chiarite tra poco, conveniamo di assumere, quale misura dello scarto tra $U(X_n, t)$ e $V(t)$, l'estremo superiore del modulo della differenza di queste funzioni (o distanza massima delle due funzioni)

$$m(X_n) = \text{estr. sup}\,|U(X_n, t) - V(t)|, \quad (-\infty < t < +\infty).$$

Allora il problema di cui ci occupiamo può essere precisato in vari modi.



SULLA DETERMINAZIONE EMPIRICA
DELLE LEGGI DI PROBABILITÀ

F. P. CANTELLI.

SUNTO. — Si mette in evidenza come in un teorema recentemente indicato dal sig. V. Glivenko « Sulla determinazione empirica delle leggi di probabilità », si possa superare la difficoltà relativa alla considerazione dei punti di discontinuità e come esso teorema possa essere convenientemente esteso.

1. La parte più interessante del teorema indicato dal Glivenko [1] riguarda la considerazione dei punti di discontinuità della legge di ripartizione limite.

In recenti conferenze tenute nel maggio scorso [2] ho perciò indicato, tra altro, una estensione del teorema accennato mettendo in rilievo come si possa superare la difficoltà relativa alla considerazione dei punti di discontinuità, basandosi sulla definizione di variabile casuale da me data fin dal 1916 [3], completando così note considerazioni [4].

« Una variabile casuale è una grandezza variabile che può assu-
« mere uno tra i valori compresi nell'intervallo (x_0, x_1) [inclusi, per
« fissare le idee, gli estremi e qualunque siano i valori reali che si
« assegnino a x_0 e x_1, essendo $x_1 > x_0$] oppure uno fra i valori esterni
« all'intervallo (x_0, x_1), secondo che si presenti uno dei due eventi
« (giudizi), escludentisi a vicenda,

[1] $E(x_0, x_1)$, $E(x_0, x_1)$

1) *Sulla determinazione empirica delle leggi di probabilità.* « Giornale dell'Istituto Italiano degli Attuari », gennaio 1933-XI, Roma. Cfr. anche l'articolo del de Finetti, pubblicato in questo numero.

2) Conferenze all'« Institut Poincaré », maggio 1933.

3) *La tendenza ad un limite nel senso del Calcolo delle probabilità.* « Rendiconti del Circolo Matematico di Palermo », 1916.

4) Cfr. ad es. P. LÉVY, *Calcul des probabilités.* Gauthier Villars, Parigi, 1925, pagina 192 e segg.

Theorem 53 (Glivenko-Cantelli). *Sind $X_1, X_2, \ldots$ reellwertig und iid mit Verteilungsfunktion F, so gilt*

$$\sup_{x \in \mathbb{R}} \left| F_n(x) - F(x) \right| \xrightarrow[n \to \infty]{f.s.} 0.$$

Beweis. Wir setzen $Y_n = \mathbb{1}_{\{X_n \leq x\}}$ und $Z_n = \mathbb{1}_{\{X_n < x\}}$. Nach dem starken Gesetz der großen Zahlen gilt

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^{n} Y_i \quad \xrightarrow{f.s.} \quad E[Y_1] = F(x) \\ F_n(x-) &= \frac{1}{n} \sum_{i=1}^{n} Z_i \quad \xrightarrow{f.s.} \quad F(x-). \end{aligned} \tag{54}$$

Es bleibt zu zeigen, dass die Konvergenz auch gleichmäßig ist. Dazu fixieren wir ein N und setzen

$$x_j^N := \inf\left\{x \in \mathbb{R} : F(x) \geq \frac{j}{N}\right\}, \quad j = 0, \ldots, N$$

und

$$R_n^N := \max_{j=1,\ldots,N-1} \left(\left| F_n(x_j^N) - F(x_j^N) \right| + \left| F_n(x_j^N-) - F(x_j^N-) \right| \right).$$

Nach Gleichung (54) gilt demnach

$$R_n^N \xrightarrow{f.s.} 0.$$

Außerdem ist für $x \in \left(x_{j-1}^N, x_j^N\right)$ $F(x_j^N-) \leq F(x) + \frac{1}{N}$ und somit

$$F_n(x) \leq F_n(x_j^N-) \leq F(x_j^N-) + R_n^N \leq F(x) + R_n^N + \frac{1}{N}$$

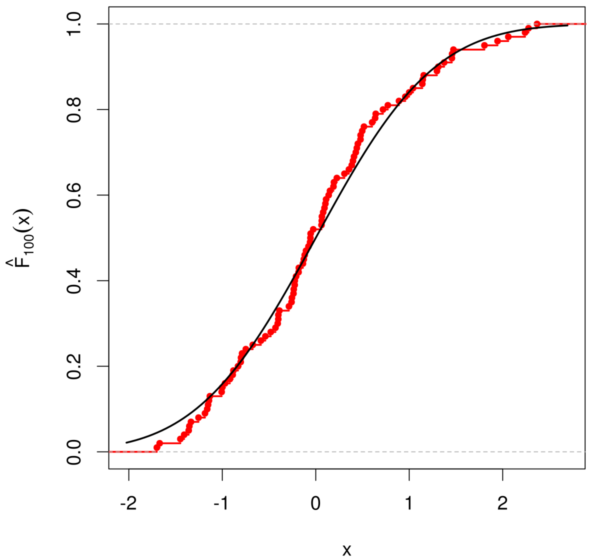
und

$$F_n(x) \geq F_n(x_{j-1}^N) \geq F(x_{j-1}^N) + R_n^N \geq F(x) - R_n^N - \frac{1}{N}.$$

Hiraus folgt, dass $\sup \left| \overline{F}_n(x) - F(x) \right| \leq \frac{1}{N} + R_n^N \xrightarrow{f.s.} \frac{1}{N}$.

Da die linke Seite nicht von N abhängt, folgt die Aussage für $N \to \infty$. $\square$

[20]



[20] Als Illustration betrachten wir dieses Bild von Wikipedia `https://de.wikipedia.org/wiki/Satz_von_Gliwenko-Cantelli`.
Es stellt großartig die Konvergenz der empirischen Verteilung gegen die wahre Verteilung dar.

Signierte Maße und der Satz von Radon-Nikodým

Wir wenden uns nur kurz diesem allgemeineren Begriff zu, eine ausführlichere Behandlung findet man im Anhang. Ein signiertes Maß darf auch negative Masse vergeben, etwa bei einer Ladungsverteilung. Sei $(\Omega, \mathscr{F})$ ein fester Messraum.

Definition 55. Eine Abbildung $\nu : \mathscr{F} \to \mathbb{R}$ heißt *signiertes Maß*, falls

(i) $\nu(\emptyset) = 0$
(ii) $\nu(\mathscr{F}) = \{\nu(F) : F \in \mathscr{F}\}$ ist entweder Teilmenge von $(-\infty, +\infty]$ oder von $[-\infty, \infty)$.
(iii) Sind $(A_n) \in \mathscr{F}$ paarweise disjunkt, so gilt

$$\nu\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \nu(A_n).$$

Beispiel 56. Hat man Maße μ_1, μ_2, wobei eines davon endlich ist, so ist $\nu = \mu_1 - \mu_2$ signiertes Maß. In der Tat hat jedes signierte Maß eine solche Gestalt und μ_1, μ_2 sind bei geeignet „minimaler" Wahl sogar eindeutig.

Wir zeigen in Satz A.40 die berühmte *Hahn-Zerlegung*: Zu jedem signierten Maß ν gibt es eine bis auf Nullmengen eindeutige disjunkte Zerlegung $\Omega = P + N$, so dass ν positive Masse auf allen messbaren Teilmengen von P vergibt und negative Masse auf aullen messbaren Teilmengen von N. Aus dieser Zerlegung ergibt sich die *Jordan-Zerlegung* eines signierten Maßes in der Form $\nu = \nu^+ + \nu^-$, siehe Satz A.42.

Für uns wird aber eine andere Darstellung viel wichtiger sein: Die Darstellung eines Maßes über seine Dichte. Dies kennen wir bereits aus der Stochastik 1, die Normalverteilung wird über ihre Dichte bezüglich des Lesbesgue-Maßes dargestellt, ebenso die Exponentialverteilung usw. Diese haben aber unterschiedliche Nullmengen, da die Exponentialverteilung die gesamte Maße auf den positiven Zahlen vergibt. Alle Lesbesgue-Nullmengen sind aber natürlich Nullmengen der beiden Verteilungen. Diese Schlüsselbeobachtung führt uns zu folgenden Definitionen und schließlich zum Satz von Radon-Nikodym.

Definition 57. Wieder betrachten wir den festen Maßraum $(\Omega, \mathscr{F})$

(i) Das Maß ν heißt *absolut stetig* bzgl. μ, falls

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \qquad \text{für alle } A \in \mathscr{F}.$$

Dann schreiben wir $\nu \ll \mu$. Gilt zusätzlich $\mu \ll \nu$, so heißen μ und ν *äquivalent* und wir schreiben $\mu \sim \nu$.

(ii) μ heißt singulär zu ν, falls es $A \in \mathscr{F}$ gibt, so dass

$$\mu(A) = 0 \quad \text{und} \quad \nu(A^c) = 0.$$

Wir schreiben $\mu \perp \nu$.

(iii) Gilt

$$\nu(A) = \int_A f d\mu,$$

für alle $A \in \mathscr{F}$, so heißt f *Dichte* von ν bzgl. μ und wir schreiben

$$\frac{d\nu}{d\mu} = f \quad \text{oder} \quad d\nu = f d\mu.$$

Sowohl die Normalverteilung als auch die Exponentialverteilung sind also absolut stetig bezüglich des Lesbesgue-Maßes. Die Normalverteilung ist sogar äquivalent zu dem Lesbesgue-Maß: Sie hat die gleichen Nullmengen. Das kann man auch daran ablesen, dass ihre Dichte strikt positiv ist.

Der Schlüssel zu dem Satz von Radon-Nikodým ist nun folgendes, einfache Lemma. Sie zeigt die Existenz einer Dichte - wie so oft wird für Existenzsäte der Satz von Hahn-Banach (oder hier in der Version des Satzes von Riesz-Frechet) verwendet.

Lemma 58. *Sind ν, ϱ endliche Maße mit $\nu \ll \varrho$, so gibt es seine messbare Funktion $h : \Omega \to [0,1]$ mit*

$$d\nu = h\, d\varrho.$$

Beweis. Für $f \in \mathcal{L}^2(\varrho)$ ist $f \mapsto \int_\Omega f d\nu$ wohldefiniert und stetig ($\mathcal{L}^2$ vollständig). Nach Satz A.35 von Riesz-Fréchet gibt es $h \in \mathcal{L}^2(\varrho)$, so dass

$$\int f d\nu = \langle f, h \rangle_\varrho = \int f h\, d\varrho.$$

Insbesondere ist dies der Fall für $f = 1_A$, $A \in \mathscr{F}$. Man sieht direkt, dass $h(\Omega) \in [0,1]$ ϱ-f.s. □

Lemma 59. *Sei μ ein Maß auf $(\Omega, \mathscr{F})$.*

(i) Ist ν σ-finit und $d\nu = f d\mu = g d\mu$, so folgt

$$f = g \qquad \mu\text{-f.s.}$$

(ii) Seien $f : \Omega \to \mathbb{R}_{\geq 0}$ und $g : \Omega \to \mathbb{R}$ messbar. Dann gilt

$$\int g \cdot (f d\mu) = \int (f \cdot g) d\mu.$$

Beweis. (i) Da ν σ-finit ist, gibt es $\Omega_1 \subseteq \Omega_2 \subseteq \cdots \subseteq \mathscr{F}$, so dass $\bigcup_{n\geq 1} \Omega_n = \Omega$ und $\nu(\Omega_n) < \infty$.

Wir setzen $A_n = \Omega_n \cap \{f < g\} \Rightarrow \int_{A_n} (f-g) d\mu = v(A_n) - \nu(A_n) = 0$, also $\mu(A_n) = 0$.

Damit ist

$$\mu(f > g) = \mu\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mu(A_n) = 0.$$

(ii) Offensichtlich ist für $g = 1_A$

$$\int g(f \cdot d\mu) = \int_A f d\mu = \int (1_A \cdot f) d\mu.$$

Die übliche Approximation von messbaren g liefert das Resultat. □

Beispiel 60. Natürlich kennen wir bereits Dichten bzgl. des Lebesgue-Maßes (Normalverteilungen etc.). Jede diskrete Verteilung hat allerdings auch eine Dichte bezüglich des Zählmaßes

$$\mu = \sum_{n\geq 1} \delta_{x_n}$$

mit dem Dirac-Maß $\delta_{x_n}(A) := \mathbb{1}_A(x_n)$.

Das einfachste Maß: Das Dirac Maß δ_x.

Satz 61 (Radon-Nikodým). *Es sei μ σ-finites Maß und ν ein signiertes Maß, so dass $\nu \ll \mu$. Dann hat ν eine (μ-f.s.) eindeutige Dichte $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ bzgl. μ.*

Beweis. Wir führen den vollständigen Beweis im Anhang, siehe Satz A.48.

Hier untersuchen wir kurz den einfacheren Fall, wo μ, ν endlich sind. Die Idee ist, Lemma 58 anzuwenden. Dazu betrachten wir $\tau = \mu + \nu$. Nach Lemma 58 existieren dann h, g so dass $d\mu = g d\tau$ und $d\nu = h d\tau$. Wir setzen $N = \{g = 0\}$, so dass $\mu(N) = 0$, also auch $\nu(N) = 0$, da $\nu \ll \mu$.

Definiere

$$f(x) := \frac{h(x)}{g(x)} \mathbb{1}_{\{x \notin N\}}.$$

Dann ist

$$\nu(A) = \nu(A \cap N^c) = \int_{A \cap N^c} h d\tau = \int_{A \cap N^c} f d\mu = \int_A f d\mu$$

und somit ist f die gesuchte Dichte. □

Als eine Anwendung des Satzes von Radon und Nikodým erhalten wir die berühmte Lesbesgue-Zerlegung in Satz A.50. Sie erlaubt die Zerlegung eines Maßes auf $\mathscr{B}(\mathbb{R})$ in einen absolut stetigen, einen diskreten und einen singulär-stetigen Teil. Er spielt auch eine wichtige Rolle in der Finanzmathematik[21], und wird darüber hinaus verwendet um die Existenz der bedingten Erwartung zu zeigen.

[21] Der erste Hauptsatz der Finanzmathematik charakterisiert arbitragefreie Märkte durch Existenz eines **äquivalenten** Martingalmaßes, so dass Maßwechsel ein wichtiges Hilfsmittel zur arbitragefreien Bewertung von Finanzprodukten sind. Mehr dazu z.B. in der Vorlesung diskrete Finanzmathematik.

Produkträume

Produkträume spielen in vielen Fällen in der Wahrscheinlichkeitstheorie eine Rolle - etwa bei einer Folge von unabhängigen Zufallsvariablen und somit bei Grenzwertsätzen. Aber auch bei stochastischen Prozessen, in welchen eine zufällige Funktion der Zeit $X_t(\omega),\ t \geq 0$ betrachtet wird. In diesem Fall sind wir bereits an überabzählbaren Indexmengen interessiert, was die zentrale Herausforderung dieses Abschnittes ist. Ebenso benötigen wir Produkträume für den bald folgenden Satz von Fubini.

Auf einem topologischen Raum $(\Omega, \mathcal{O})$ und für ein $A \subseteq \Omega$ definieren wir das Innere A^o und den Abschluss $\overline{A}$ von A durch

$$A^o := \bigcup \{O \subseteq A : O \in \mathcal{O}\}, \quad \overline{A} := \bigcap \{F \supseteq A : F^c \in \mathcal{O}\}$$

Eine wichtige Annahme wird die Beschränkung auf polnische Räume sein[22].

[22] Die zentralen Grenzwertsätze der Stochastik werden auf **polnischen** Räumen gelten.

Definition 62. Ein topologischer Raum $(\Omega, \mathcal{O})$ heißt **polnisch**, falls

(i) es eine abzählbare Menge $\Omega' \subseteq \Omega$ gibt, so dass $\overline{\Omega'} = \Omega$,
(ii) er vollständig metrisierbar ist.

Ein topologischer Raum heißt *vollständig metrisierbar*, falls es eine Metrik gibt, die die Topologie erzeugt und bzgl. der Ω vollständig (jede Cauchy-Folge konvergiert) ist.

Beispiel 63. (i) $(\mathscr{C}[0,1], \|\cdot\|_\infty)$ ist polnisch
(ii) jeder **kompakte** metrische Raum ist polnisch
(iii) $(\mathscr{C}[0,\infty), d)$ mit

$$d(f,g) = \sum_{k \geq 1} \frac{1}{2^k} \frac{d_k(f,g)}{1 + d_k(f,g)} \qquad \text{mit } d_k = \sup_{[0,k]} |f - g|$$

ist polnisch (Topologie der gleichmäßigen Konvergenz auf Kompakta).
(iv) Ein nicht polnischer Raum ist z.B. der diskrete Raum - der Raum in der jede Teilmenge offen ist. Ist er nicht endlich, ist er nicht separabel.

Für eine Familie $(\Omega_i)_{i\in I}$ von Mengen definieren wir den **Produktraum** durch
$$\bigtimes_{i\in I} \Omega_i := \left\{(\omega_i)_{i\in I} : \omega_i \in \Omega_i\right\}.$$
Wir schreiben hierfür im Folgenden kurz $\Omega_I := \times_{i\in I}\Omega_i$. Für Teilmengen $H \subseteq J \subseteq I$ definieren wir die zugehörigen **Projektionen** $\pi_{J,H} : \Omega_J \to \Omega_H$ durch
$$\pi_{J,H}((\omega_j)_{j\in J}) = (\omega_h)_{h\in H}.$$

Außerdem schreiben wir $\pi_H = \pi_{I,H}$ und $\pi_i = \pi_{\{i\}}$, $i \in I$. Ausgehend von den Borel-σ-Algebren auf topologischen Räumen können wir auch eine Topologie auf dem Produktraum definieren.

Definition 64. Ist $(\Omega_i, \mathcal{O}_i)_{i\in I}$ eine Familie von topologischen Räumen, so heißt die von
$$\left\{ \bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_j \in \mathcal{O}_j,\ j \in J\right\}$$
erzeugte Topologie die *Produkttopologie* auf $\Omega = \Omega_I$.

Bezüglich der Produkttopologie $\mathcal{O}$ sind alle Projektionen π_i, $i \in I$ stetig:
$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{j\in I\setminus\{i\}} \Omega_j \in \mathcal{O}, \qquad A_i \in \mathcal{O}_i.$$

Besonders schöne Eigenschaften erhält man unter Abzählbarkeit.

Satz 65. *Sei $(\Omega_i, \mathcal{O}_i)_{i\in\mathbb{N}}$ eine Familie polnischer Räume. Dann ist der Produktraum Ω versehen mit der Produkttopologie $\mathcal{O}$ wieder polnisch.*

Beweis. Da Ω_i separabel ist, gibt es abzählbare $\Omega_i' \subseteq \Omega_i$, so dass $\overline{\Omega_i'} = \Omega_i$ für alle $i \in I$. Weiterhin bezeichne τ_i die vollständige Metrik, die $\mathcal{O}_i$ erzeugt (über $\{B_\varepsilon(\omega) = \{\omega' \in \Omega : d(\omega,\omega') < \varepsilon\} : \varepsilon > 0, \omega \in \Omega\}$).

Für $\omega, \omega' \in \Omega = \times_{i\in\mathbb{N}}\Omega_i$ wird durch
$$\tau(\omega,\omega') = \sum_{i\geq 1} \frac{1}{2^i}(\tau_i(\omega_i,\omega_i') \wedge 1)$$
eine vollständige Metrik auf Ω definiert, die die Produkt-Topologie $\mathcal{O}$ erzeugt.

Für $\omega' \in \times_{i\in\mathbb{N}}\Omega_i'$ ist
$$B_{\omega'} := \left\{\omega \in \times_{i\in\mathbb{N}}\Omega_i' : \omega_i \neq \omega_i' \text{ für endlich viele } i \in \mathbb{N}\right\}$$

abzählbar und dicht in Ω und die Behauptung folgt. $\square$

Definition 66. Für eine Familie $(\Omega_i, \mathscr{F}_i)_{i\in I}$ von Maßräumen heißt die σ-Algebra

$$\bigotimes_{i\in I} \mathscr{F}_i := \sigma\Big(\{A_J \times \Omega_{I\setminus J} : J \subseteq I, I \text{ endlich}, A_j \in \mathscr{F}_j, j \in J\}\Big)$$

die Produkt-σ-Algebra auf $\Omega = \Omega_I$. Ist $\mathscr{F}_i = \mathscr{F}_1 =: \mathscr{F}$, so schreiben wir auch $\mathscr{F}^I = \otimes_{i\in I}\mathscr{F}$.

$A_J \times \Omega_{I\setminus J} = \times_{j\in J} A_j \times \times_{i\in I\setminus J}\Omega_i$

Wieder sind die Projektionen verträglich mit der Konstruktion: Da

$$\pi_i^{-1}(A_i) = A_i \times \Omega_{I\setminus\{i\}} \in \bigotimes_{i\in I} \mathscr{F}_i,$$

ist π_i messbar bzgl. $\otimes_{i\in I}\mathscr{F}_i$. Außerdem ist

$$\bigotimes_{i\in I} \mathscr{F}_i = \sigma\Big(\Big\{A_i \times \Omega_{J\setminus\{i\}} : A_i \in \mathscr{F}_i, i \in I\Big\}\Big).$$

Unter geeigneten Voraussetzungen ist die Borel-σ-Algebra der Produkttopologie gleich der Produkt-σ-Algebren der einzelnen Borel-σ-Algebren. Abzählbarkeit ist eine solche (vgl. Elstrodt, Kapitel III.5).

Lemma 67. *Ist I abzählbar, $\Omega = \Omega_I$, $\mathcal{O}$ die Produktopologie auf Ω und $(\Omega, \mathcal{O})$ polnisch, so gilt*

$$\mathscr{B}(\mathcal{O}) = \otimes_{i\in I}\mathscr{B}(\mathcal{O}_i).$$

Beweis: Wir nutzen folgendes, einfaches topologisches Resultat: Ist $(\Omega, \mathcal{O})$ separabel, so hat $\mathcal{O}$ eine abzählbare Basis $\mathcal{B}$, d.h.

$$\mathcal{O} = \Big\{\bigcup_{i\in I} B_i : B_i \in \mathcal{B},\ I \subset \mathbb{N}\Big\}$$

Da alle $(\Omega_i, \mathcal{O}_i), i \in I$, separabel sind, haben sie jeweils abzählbare Basen $\mathcal{B}_i$. Dann ist

$$\mathcal{B}' := \Big\{A_J \times \Omega_{I\setminus J} : J \subseteq I, J \text{ endlich}, A_i \in \mathcal{B}_i\Big\}$$

eine abzählbare Basis von $(\Omega, \mathcal{O})$ und somit $\sigma(\mathcal{B}') = \mathscr{B}(\Omega)$. Außerdem ist $\mathcal{B}' \subseteq \otimes_{i\in J}\mathscr{B}(\Omega_i)$, also $\mathscr{B}(\Omega) \subseteq \otimes_{i\in I}\mathscr{B}(\Omega_i)$.

Umgekehrt ist für $A_i \in \mathcal{O}_i$

$$A_i \times \Omega_{I\setminus\{i\}} \in \sigma\Big(\{A_i \times \Omega_{I\setminus\{i\}} : A_i \in \mathcal{O}_i\}\Big) \subseteq \mathscr{B}(\Omega). \qquad \square$$

Der Satz von Fubini

Produkträume sind ein wesentliches Hilfsmittel zum Studium der Verteilung von einer Familie von Zufallsvariablen. Ein zweites wesentliches Hilfsmittel sind Übergangskerne. Dies erhält man durch bedingen:

$$P(X \in A, Y \in B) = E\big[P(X \in A|Y)\mathbb{1}_{\{Y \in B\}}\big],$$

so dass für die gemeinsame Verteilung einmal die Verteilung von Y benötigt wird und die Verteilung von X bedingt auf Y. Die bedingte Verteilung ist eine zufällige Verteilung, da sie von Y abhängt, und der folgende Abschnitt erklärt präzise wie wir damit umgehen.

Ein wichtiges Hilfsmittel für Maße auf Produkträumen (und für unser obiges Beispiel) sind **Übergangskerne**. Sie beschreiben in einem dynamischen Kontext Übergangswahrscheinlichkeiten von so genannten Markov-Prozessen, wie wir gleich in einem Beispiel sehen werden.

Definition 68. Seien $(\Omega_1, \mathscr{F}_1)$ und $(\Omega_2, \mathscr{F}_2)$ Messräume. Eine Abbildung $\kappa : \Omega_1 \times \mathscr{F}_2 \to \mathbb{R}_{\geq 0}$ heißt *Übergangskern*, falls

(i) für alle $\omega_1 \in \Omega_1$ ist $\kappa(\omega_1, \cdot)$ ein Maß auf $(\Omega_2, \mathscr{F}_2)$
(ii) für alle $A_2 \in \mathscr{F}_2$ ist $\kappa(\cdot, A_2)$ $\mathscr{F}_1$-messbar.

Ein Übergangskern heißt σ-finit, falls es $A_1, A_2, \ldots \uparrow \Omega_2$ gibt mit $\sup_{\Omega_1} \kappa(\omega_1, A_n) < \infty$ für alle $n = 1, 2, \ldots$

Er heißt *stochastischer Kern* , falls

$$\kappa(\omega_1, \Omega_2) = 1 \qquad \text{für alle } \omega_1 \in \Omega_1.$$

Für einen stochastischen Kern verwenden wir auch oft die Bezeichnung Markovscher Kern oder Übergangswahrscheinlichkeit.

Beispiel 69 (Markov-Kette)**.** Sei $\Omega = \{\omega_1, \ldots, \omega_n\}$ mit Wahrscheinlichkeiten $(p_{ij})_{1 \leq i,j \leq n}$, so dass mit $\sum_{j=1}^n p_{ij} = 1 \forall i$. Dann ist

$$\kappa(\omega_i, \cdot) := \sum_{j=1}^n p_{ij} \delta_{\omega_j}$$

ein stochastischer Kern von $(\Omega, \mathscr{P}(\Omega))$ nach $(\Omega, \mathscr{P}(\Omega))$. Die (p_{ij}) bilden die Übergangswahrscheinlichkeiten einer homogenen **Markov-Kette**.

Ein typisches Beispiel kann man sich so vorstellen. Betrachten wir als Zustandsraum $\Omega = \mathbb{Z}$. Ausgehend von einem Punkt x geht die Markovkette entweder einen Schritt nach oben, oder einen nach unten, so dass der Übergangskern

$$\kappa(x,\cdot) = p\delta_{x+1} + (1-p)\delta_{x-1}$$

ist. Die Frage ist nun: Reicht die Angabe dieser Übergangswahrscheinlichkeiten um die Verteilung des Markov Prozesses genau zu beschreiben.

Die Messbarkeit von Schnitten

Lemma 70. *Sei κ ein σ-finiter Übergangskern und $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_{\geq 0}$ $\mathscr{F}_1 \otimes \mathscr{F}_2$-messbar. Dann ist*

$$\omega_1 \mapsto \int f(\omega_1,\omega_2)\kappa(\omega_1, d\omega_2)$$

$\mathscr{F}_1$-messbar.

Beweis (Skizze). Zunächst nehmen wir $\kappa(\omega_1,\Omega_2) < \infty$ an. (Dann Erweiterung mittels $A_1, A_2, \ldots \uparrow \Omega_2$.) Wir betrachten

$$\mathscr{D} = \Big\{A \in \mathscr{F}_1 \otimes \mathscr{F}_2 : \omega_1 \mapsto \int \mathbb{1}_A(\omega_2)\kappa(\omega_1, d\omega_2) \text{ ist } \mathscr{F}_1\text{-messbar}\Big\}$$

Dies ist ein durchschnittstabiles Dynkin-System mit $\mathscr{H} = \{A_1 \times A_2, A_i \in \mathscr{F}_i\} \subseteq \mathscr{D}$. Nun gilt $\sigma(\mathscr{H}) = \mathscr{F}_1 \otimes \mathscr{F}_2$, also $\sigma(\mathscr{H}) = \mathscr{F}_1 \times \mathscr{F}_2 \subseteq \mathscr{D} \subseteq \mathscr{F}_1 \otimes \mathscr{F}_2$ und die Aussage folgt für $f = \mathbb{1}_A$, $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$.

Wir erhalten die Aussage für Treppenfunktionen und mit monotoner Konvergenz für alle messbaren $f \geq 0$. □

Der Satz von Ionescu-Tulcea zeigt, dass man endlich viele Übergangskerne eindeutig zu einem Maß fortsetzen kann und ist somit eine typische Anwendung des Maßfortsetzungssatzes, Satz A.22.

Theorem 71 (Ionescu-Tulcea). *Seien $(\Omega_i, \mathscr{F}_i)$, $i = 0, \ldots, n$ Messräume, μ ein σ-finites Maß auf $(\Omega_0, \mathscr{F}_0)$ und κ_i σ-finite Übergangskerne von $(\times_{j=0}^{i-1}\Omega_j, \otimes_{j=0}^{i-1}\mathscr{F}_j)$ nach $(\Omega_i, \mathscr{F}_i)$, $i = 1, \ldots, n$.*

Dann gibt es genau ein σ-finites Maß $\kappa = \mu \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$ auf $(\times_{j=0}^{n}\Omega_j, \otimes_{j=0}^{n}\mathscr{F}_j)$, so dass

$$\begin{aligned}\kappa(A_0 \times \cdots \times A_n) \\ &= \int_{A_0} \cdots \int_{A_n} \kappa_n(\omega_0, \ldots, \omega_{n-1}, d\omega_n) \cdots \kappa_1(\omega_0, d\omega_1)\mu(d\omega_0). \quad (72)\end{aligned}$$

Beweis. Wir beweisen die Aussage für $n = 1$, der allgemeine Fall folgt dann per Induktion. Wir nutzen den Maßfortsetzungssatz, Satz A.22: Wir starten mit dem durchschnittstabilen Halbring

$$\mathscr{H} = \left\{ \times_{i=0}^{n} A_i : A_i \in \mathscr{F}_i \right\}$$

und betrachten die durch (72) definierte Mengenfunktion κ auf $\mathscr{H}$.

(i) *κ ist σ-finit:* Seien $(\Omega_n^i) \subseteq \mathscr{F}_i$ so dass $(\Omega_n^i) \uparrow \Omega^i$, $i = 0, 1$ und $\mu(\Omega_n^i) < \infty$ und $\sup_{\omega^0 \in \Omega^0} \kappa_1(\omega_0, \Omega_n^i) =: C_n < \infty$.

$$\Rightarrow \qquad \kappa(\Omega_n^0 \times \Omega_n^1) = \int_{\Omega_n^0} \int_{\Omega_n^0} \kappa(\omega^0, d\omega^1) \mu(d\omega^0)$$
$$\leq C_n \mu(\Omega_n^1) < \infty.$$

Da $\Omega_n^0 \times \Omega_n^1 \uparrow \Omega_0 \times \Omega_1$ ist κ σ-finit.

(ii) *κ ist σ-additiv:* Für $A_1, A_2, \dots \in \mathscr{H}$ p. d. mit $A = \sum_{i=1}^{\infty} A_i \in \mathscr{H}$ ist

$$\begin{aligned}
\kappa(A) &= \iint \mathbb{1}_A(\omega_0, \omega_1) \kappa_1(\omega_0, d\omega_1) \mu(d\omega_0) \\
&= \int \sum_{n \geq 1} \int \mathbb{1}_{A_n}(\omega_0, \omega_1) \kappa_1(\omega_0, d\omega_1) \\
&\overset{(*)}{=} \sum_{n \geq 1} \iint \mathbb{1}_{A_n}(\omega_0, \omega_1) \kappa_1(\omega_0, d\omega_1) \mu(d\omega_0) = \sum_{n \geq 1} \kappa(A_n),
\end{aligned}$$

wobei (*) aus monotoner Konvergenz folgt, da sie Summanden jeweils nicht-negativ sind. Mit dem Maßfortsetzungssatz erhalten wir die Behauptung. □

Bemerkung 73. Mit etwas mehr Aufwand kann man die Aussage auch für $n = \infty$ beweisen, siehe [23].

[23] V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1991

Theorem 74 (Satz von Fubini)**.** *Seien $(\Omega_i, \mathscr{F}_i)$, μ_0, κ_i für $i = 1, \dots, n$ und $\kappa = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$ wie in Theorem 71 gegeben und die Funktion $f : \times_{i=0}^{n} \Omega_i \to \mathbb{R}_{\geq 0}$ messbar bzgl. $\otimes_{i=0}^{n} \mathscr{F}_i$. Dann gilt*

$$\int f d\kappa \tag{75}$$
$$= \int \Big(\cdots \Big(\int f(\omega_0, \dots, \omega_n) \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) \Big) \cdots \kappa_1(\omega_0, d\omega_1) \Big) \mu(d\omega_0).$$

Die Aussage gilt auch für beliebig messbare f mit $\int |f| d\kappa < \infty$.

Beweis. Wir definieren die Abbildung $I : \otimes_{i=0}^{n} \mathscr{F}_i \mapsto \bar{\mathbb{R}}$ durch $I(A) := \int_A d\kappa$. Für alle $A \in \left\{ \times_{i=0}^{n} A_i : A_i \in \mathscr{F}_i \right\}$ gilt nach Gleichung (72), dass $I(A) = \kappa(A)$ und somit (75) für $f = \mathbb{1}_A$.

Wegen Linearität des Integrals und mit Hilfe des Satzes von der monotonen Konvergenz folgt (75) für alle nicht-negativen, messbaren Funktionen.

Die Messbarkeit der einzelnen Integrale folgt hierbei nach Lemma 70. □

Bemerkung 76. Die Aussage für Produktmaße $\mu_0 \otimes \cdots \otimes \mu_n$ erhält man als einfachen Spezialfall!

Beispiel 77 (Mehrdimensionales Lebesgue-Maß)**.** Mit $\lambda_d = \otimes_{i=1}^d \lambda$ bezeichnen wir das Lebesgue-Maß auf dem $\mathbb{R}^d$

$$f(x,y) = \frac{xy}{(x^2+y^2)^2} \Rightarrow \int f(x,y)dy = 0 \quad \forall x$$

$$\Rightarrow \qquad \iint f(x,y)\lambda(dy)\lambda(dx) = \iint f(x,y)dx)\lambda(dy) = 0.$$

Allerdings ist $|f|$ nicht bzgl. λ_2 integrierbar!

Aus Gleichheit und Endlichkeit der Mehrfachintegrale kann man also nicht auf die Integrierbarkeit des Integranden schließen.

Die Faltung

Betrachten wir zwei unabhängige Zufallsvariable X und Y, so heißt die Verteilung von $X+Y$ die **Faltung**.

Die Faltung ist demnach die Verteilung $P(X+Y \leq z)$ für alle $z \in \mathbb{R}$.

Wir erinnern noch kurz an die Definition A.25 wo wir das Bildmaß (oder den push-forward) eines Maßes unter einer Abbildung eingeführt haben. Dies kann man nutzen um etwa die Faltung präzise einzuführen.

Definition 78. Seien $\mu_1, \ldots, \mu_n$ σ-finite Maße auf $\mathscr{B}(\mathbb{R})$, $\mu = \otimes_{i=1}^n \mu_i$ deren Produktmaß und $S(x) = x_1 + \cdots + x_n$, $x \in \mathbb{R}^n$. Dann heißt

$$S_*\mu := \mu_1 * \cdots * \mu_n$$

die *Faltung* der Maße $\mu_1, \ldots, \mu_n$.

Aufgabe 1. Haben die Maße μ_i Dichten f_i bzgl. λ, $i = 1,2$ und setzt man

$$f(t) := \int f_1(s) f_2(t-s)\lambda(ds),$$

so gilt

$$\mu_1 * \mu_2 = f_{\mu * \nu} \cdot \lambda. \tag{Fubini}$$

Betrachte $\quad P(X+Y \leq x) = P(X \leq x - Y)$!

Aufgabe 2. Die Faltung zweier Normalverteilungen ist wieder normalverteilt.

Der Existenzsatz von Kolmogorov

Der Existenzsatz von Kolmogorov ist der zentrale Grund, warum wir im Folgenden stets polnische Räume betrachten werden.

Bisher haben wir endliche bzw. abzählbare Produkträume betrachtet, was für unsere Anwendungen nicht ausreichen wird. Wenn wir an die Brownsche Bewegung denken - also an eine stetige Funktion, so werden wir auch an allgemeineren Situationen interessiert sein.

Der Satz von Kolmogorov erlaubt die Erweiterungen auf größere Räume, kommt allerdings nicht ohne die Voraussetzung aus, dass diese *polnisch* sind.

Wir erinnern daran, dass ein topologischer Raum $(\Omega, \mathcal{O})$ *polnisch* heißt, falls er vollständig metrisierbar ist und separabel.

- vollständig metrisierbar: Es gibt eine Metrik r so dass $\mathcal{O}$ von r erzeugt wird, d.h. für jedes $A \in \mathcal{O}$ gibt es ein B_ε, so dass $B_\varepsilon \subseteq A$ ist. ($B_\varepsilon(\omega) = \{\omega' \in \Omega : r(\omega, \omega') < \varepsilon\}$) <u>und</u> dass (Ω, r) vollständig ist (jede Cauchy-Folge konvergiert)

- separabel: Es gibt eine abzählbare Teilmenge Ω', so dass

$$\overline{\Omega}' = \bigcap \{A \supseteq \Omega' : A^c \in \mathcal{O}\} = \Omega.$$

Beispiel 79. (i) $\mathbb{R}^d = \overline{\mathbb{Q}}^d$ ist separabel und vollständig

(ii) Sei $K \subseteq \mathbb{R}$ kompakt $\Rightarrow (C(K), \|\cdot\|_\infty)$ ist **polnisch** (Weierstraß'scher Approximationssatz – Polynome)

Wie bisher bezeichnen wir $\mathscr{F}^J = \otimes_{j \in J}$. Wir verwenden im Folgenden die praktische Konvention

$$J \Subset I := \{J \subseteq I : \ J \text{ endlich } \}.$$

In diesem Sinne ist etwa $(P_J)_{J \Subset I} = (P_J : J \subseteq I, J \text{ endlich})$ zu verstehen.

Definition 80. Sei $(\Omega, \mathscr{F})$ ein Maßraum und I eine beliebige Indexmenge. Eine Familie von Wahrscheinlichkeitsmaßen $(P_J)_{J \Subset I}$ heißt *projektive Familie*, falls P_J Wahrscheinlichkeitsmaß auf $(\Omega^J, \mathscr{F}^J)$ ist und

$$P_H = \pi_{J,H*} P_J$$

für alle $H \subseteq J \subseteq I$, J endlich.

Die Idee hierbei ist, Produktmaße wie im Satz von Ionescu-Tulcea auf unendliche Produkte zu erweitern. Wir betrachten die Familien $P_n = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$, $n \geq 1$. Natürlicherweise gilt für die endlichen Produkte, dass

$$P_i(A_1 \times \Omega_2 \times \cdots \times \Omega_i) = P_j(A_1 \times \Omega_2 \times \cdots \times \Omega_j).$$

Eine projektive Familie besitzt ebendiese Konsistenzeigenschaft, ohne dass wir verlangen, dass die Familie als Produktmaß erzeugt wurde.

Definition 81. Existiert für eine projektive Familie $(P_J)_{J \Subset I}$ ein Wahrscheinlichkeitsmaß P_I auf $\mathscr{F}^I = \otimes_{i \in I} \mathscr{F}_i$ mit $P_J = \pi_{J*} P_I$ für alle $J \subseteq I$, J endlich, so heißt P_I *projektiver Limes* der Familie P. Wir schreiben

$$P_I = \varprojlim_{J \Subset I} P_J.$$

In mindestens zwei Situationen spielen projektive Familien eine große Rolle.

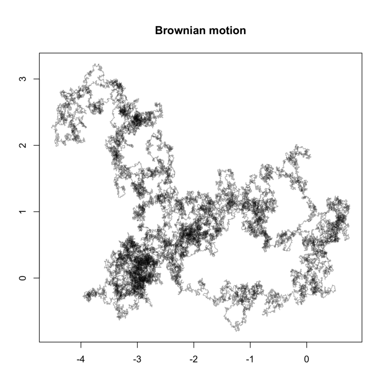
Beispiel 82. Sei $(\Omega, \mathscr{F}, P)$ ein Wahrscheinlichkeitsraum und I eine unendliche Indexmenge. In Theorem 71 haben wir für jedes $J \Subset I$ das Produktmaß $P^{\otimes J}$ auf $\mathscr{F}^J$ definiert.

Zentrale Beobachtung ist, dass die Familie $(P^{\otimes J})_{J \Subset I}$ projektiv ist: Ist nämlich $H \subseteq J \Subset I$, so ist für $A_i \in \mathscr{F}, i \in H$,

$$\begin{aligned}
(\pi_{J,H})_* P^{\otimes J}\big(\times_{i \in H} A_i\big) &= P^{\otimes J}\big((\pi_{J,H})^{-1}\big(\times_{i \in H} A_i\big)\big) \\
&= P^{\otimes J}\big(\times_{i \in H} A_i \times \times_{i \in J \setminus H} \Omega\big) \\
&= \prod_{i \in H} P(A_i) \cdot \prod_{i \in J \setminus H} P(\Omega) \\
&= \prod_{i \in H} P(A_i) \\
&= P^{\otimes H}\big(\times_{i \in H} A_i\big).
\end{aligned}$$

Allerdings haben wir noch nicht gezeigt, wann es den projektiven Limes von $(P^{\otimes J})_{J \Subset I}$ gibt. Diesen würden wir dann das unendliche Produktmaß $P^{\otimes I}$ nennen.

Beispiel 83 (Brownsche Bewegung)**.** Die Brownsche Bewegung ist ein stetiger Gaußprozess mit unabhängigen und identisch verteilten Zuwächsen. Wir verwenden folgende, diskrete Approximation:



$$W(t) = \sum_{i: t_i \le t} \xi_i$$

mit $t_i = i\Delta$ und i.i.d. $\xi \sim \mathscr{N}(0, \Delta)$.

Wir simulieren dies mit folgendem R-Code (damit wurde auch das Titel-Bild erstellt - erzeugen Sie Ihre Lieblings-Brownsche Bewegung und schicken Sie mir eine Version zu!)[24]:

[24] In korrigierter Version - vielen Dank an Eni Schwindt.

```
# Simulate Brownian motion
bm <- function (n,delta)
{
    # take the cumulative sum of normally distributed random variables
    cumsum(rnorm(n,0,sd=sqrt(delta)))
}

## two-dimensional Brownian motion

datax = c(0,bm(100,0.0001))
datay = c(0,bm(100,0.0001))

# now we plot it
plot(datax,datay,type="l",col = rgb(0, 0, 0, 0.25))

# for an artistic output without boundaries and text:
plot(datax,datay,type="l",col = rgb(0, 0, 0, 0.25),
     type="l",main="",xlab="",ylab="",axes=FALSE)
```

Beispiel 84. Sei $(\Omega, \mathscr{F}, P)$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, $(\Omega', \mathscr{F}')$ ein Messraum und $X_i : \Omega \to \Omega', i \in I$ eine Familie von Zufallsvariable. Wir werden eine solche Familie $X := (X_i)_{i \in I}$ einen *stochastischen Prozess* nennen.

So ist z.B. $W = (W_t)_{t \geq 0}$ - die Brownsche Bewegung - ein stochastischer Prozess, d.h. eine Familie von Zufallsvariablen.

Dann ist formal gesehen $X : \Omega \to \Omega'^I$ mit $X(\omega) = (X_i(\omega))_{i \in I}$. Man kann sich nun fragen, ob die Verteilung von X (d.h. das Bildmaß $X_* P$) als Verteilung auf $\mathscr{F}'^I$ existiert.

Hierzu sei bemerkt, dass $P'_J := ((X_j)_{j \in J})_* P, J \Subset I$ eine projektive Familie ist. Ist nämlich $H \subset J \Subset I$ und $\widetilde{A}_i \in \mathscr{F}', i \in H$, dann ist

$$\begin{aligned}
(\pi_{J,H})_* P'_J\big(\times_{j \in H} \widetilde{A}_j\big) &= P'_J\big((\pi_{J,H})^{-1} \times_{j \in H} \widetilde{A}_j\big) \\
&= P'_J\big(\times_{j \in H} \widetilde{A}_j \times \times_{j \in J \setminus H} \Omega'\big) \\
&= P\big(X_j \in \widetilde{A}_j, j \in H \text{ und } X_j \in \Omega', j \in J \setminus H\big) \\
&= P\big(X_j \in \widetilde{A}_j, j \in H\big) \\
&= P'_H\big(\times_{j \in H} \widetilde{A}_j\big).
\end{aligned}$$

Wie im Folgenden gezeigt wird, gibt es die Verteilung $X_* P$ (was dann der projektive Limes der $(P'_J)_{J \Subset I}$ ist) zumindest dann, wenn $\mathscr{F}'$ die Borel'sche σ-Algebra eines polnischen Raumes ist.

Bemerkung 85 (Eindeutigkeit des projektiven Limes)**.** Seien P_I und P'_I zwei projektive Limiten, so ist für $A := \times_{i \in J} A_i \times \times_{i \in I \setminus J} \Omega_i \in \mathscr{H}$ mit $\mathscr{H}$ aus Lemma A.58 und $J \Subset I$

Zu jeder projektiven Familie gibt es höchstens *einen* projektiven Limes.

$$P_I(A) = P_J\big(\times_{i \in J} A_i\big) = P'_J\big(\times_{i \in J} A_i\big) = P'_I(A).$$

Damit stimmen P_I und P'_I auf dem schnittstabilen Erzeuger überein und nach Satz A.16 gilt $P_I = P'_I$.

Das folgende, berühmte Theorem, welches unabhängig voneinander von P.J. Daniell (welcher abzählbares I betrachtet hat [25]) und A.N. Kolmogorov gezeigt wurde besagt nun, dass auf **polnischen** Räumen ein solcher Limes stets existiert.

[25] Percy John Daniell. Functions of limited variation in an infinite number of dimension. *Annals of Mathematics*, pages 30–38, 1919

Theorem 86 (Existenz von Prozessen)**.** *Sei* $(\Omega, \mathcal{O})$ *ein* polnischer *Raum,* $\mathscr{F} = \mathscr{B}(\mathcal{O})$ *und* $(P_J)_{J \Subset I}$ *eine projektive Familie von Wahrscheinlichkeitsmaßen auf* $\mathscr{F}$*. Dann existiert der projektive Limes*

$$P_I := \varprojlim_{J \Subset I} P_J.$$

Beweis. Sei $\mathscr{H}$ wie in Lemma A.58 und μ eine additive Mengenfunktion auf dem Halbring $\mathscr{H}$, definiert durch

$$\mu\Big(\times_{j \in J} A_j \times \times_{i \in I \setminus J} \Omega \Big) := P_J \Big(\times_{j \in J} A_j \Big).$$

Damit und μ ein wohldefinierter Inhalt auf dem Halbring $\mathscr{H}$.

Für die σ-Additivität von μ muss man etwas weiter ausholen und sich Kompaktheitsargumente zu nutze machen, die wir hier zunächst ausklammern. (Für einen leicht andern Beweis siehe Klenke [26]).

[26] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2013

Weiterhin ist $\mu(\Omega^I) = 1$, also lässt μ mit dem Maßfortsetzungssatz, Theorem A.22, in eindeutiger Weise auf ein Maß P_I auf $\sigma(\mathscr{H}) = \mathcal{F}^I$ fortsetzen. Dies muss wegen Eindeutigkeit der projektive Limes von $(P_J)_{J \Subset I}$ sein. $\square$

Schwache Konvergenz

Schwache Konvergenz

Wir betrachten den metrischen Raum (E, d) und bezeichnen

$$\begin{aligned}\mathscr{P}(E) &= \{\mu : \mu \text{ Wahrscheinlichkeitsmaß auf } \mathscr{B}(E)\} \\ \mathscr{P}_{\leq 1}(E) &= \{\mu : \mu \text{ Maß auf } \mathscr{B}(E) \text{ mit } \mu(E) \leq 1\}.\end{aligned}$$

Als Testfunktionen verwenden wir $C_b(E)$: die Menge der stetigen, beschränkten, reellwertigen Funktionen auf E, oder $C_c(E)$: die Menge der stetigen, reellwertigen Funktionen mit kompaktem Träger.

Definition 87. Die Folge $P_1, P_2, \dots \in \mathscr{P}(E)$ konvergiert *schwach* gegen $P \in \mathscr{P}(E)$, falls

$$\int f dP_n \underset{n\to\infty}{\longrightarrow} \int f dP$$

für alle $f \in C_b(E)$. Wir schreiben dann $P_n \Rightarrow P$.

Die vage Konvergenz definieren wir für Maße mit der Masse höchstens eins, weswegen die Bezeichnung μ verwendet wird (an Stelle von P - was typischerweise nur für Wahrscheinlichkeitsmaße verwendet wird).

Definition 88. Seien $\mu_1, \mu_2, \dots \subseteq \mathscr{P}_{\leq 1}(E)$ und μ ein Maß auf $\mathscr{B}(E)$. Gilt

$$\int f d\mu_n \to \int f d\mu$$

für alle $f \in C_c(E)$, so konvergiert die Folge (μ_n) *vage* gegen μ, und wir schreiben $\mu_n \Rightarrow_{\mathrm{v}} \mu$.

Definition 89. Seien $X, X_1, X_2, \dots$ Zufallsvariable auf $(\Omega, \mathscr{F}, P)$, $(\Omega_1, \mathscr{F}_1, P_1)$,$(\Omega_2, \mathscr{F}_2, P_2)$, … mit Werten in E. Wir sagen $X_1, X_2, \dots$ *konvergieren in Verteilung* gegen X, falls $(X_n)_* P_n \Rightarrow X_* P$ und schreiben hierfür

$$X_n \Rightarrow X.$$

Oft verwendet man für die Verteilungskonvergenz auch folgende Notation:

$$X_n \xrightarrow{\mathscr{L}} X$$

wobei $\mathscr{L}$ hier Konvergenz "in law" - also in Verteilung - bedeutet.

Bemerkung 90. • Damit gilt $X_n \Rightarrow X \Leftrightarrow P_n(f(X_n) \in \cdot) \to P(f(X) \in \cdot)\ \forall f \in C_b(E)$.

- Da $1 \in C_b(E)$ ist der schwache Limes von Wahrscheinlichkeitsmaßen immer ein Wahrscheinlichkeitsmaß.

  Allerdings ist $1 \notin C_c(E)$, so dass dies für die vage Konvergenz nicht gelten muss.

- Die schwache Topologie ist die schwächste Topologie bzgl. der $P \to \int f dP$ für alle $f \in C_b(E)$ stetig ist.

Beispiel 91. (i) Seien $x, x_1, x_2, \ldots \in \mathbb{R}$ mit $x_n \to x$. Setze $P = \delta_x$, $P_n = \delta_{x_n}$. Dann gilt $P_n \Rightarrow P$, da

$$\int f dP_n = f(x_n) \to f(x) = \int f dP$$

für alle $f \in C_b(R)$.

(ii) Gilt $x_n = n$, so folgt $P_n \Rightarrow_v 0$, allerdings nicht $P_n \Rightarrow 0$, da 0 kein Wahrscheinlichkeitsmaß ist.

(iii) Sind $X, X_1, X_2, \ldots$ identisch verteilt, so folgt $X_n \Rightarrow X$, allerdings gilt weder f. s. noch stochastische Konvergenz.

(iv) Der zentrale Grenwertsatz.

Lemma 92 (Eindeutigkeit des Grenzwertes)**.** *Gilt $\mu_n \Rightarrow \mu$ und $\mu_n \Rightarrow \nu$, so folgt $\mu = \nu$.*

Beweis. Nach Satz A.16[27] reicht es, $\mu(A) = \nu(A)$ für alle abgeschlossenen $A \subseteq E$ zu zeigen, da die Menge der abgeschlossenen Mengen ein durchschnittsstabiler Erzeuger von $\mathscr{B}(E)$ ist. Setze

$$d(x, A) = \inf_{y \in A} d(x, y)$$
$$f_m(x) := \big(1 - m \cdot d(x, A)\big)^+.$$

Dann gilt, dass f_m monoton gegen $\mathbb{1}_A$ konvergiert, also

$$\mu(A) = \lim_{m\to\infty} \int f_m d\mu = \lim_{m\to\infty} \lim_{n\to\infty} \int f_m d\mu_n = \lim \int f_m d\nu = \nu(A)$$

wegen majorisierter Konvergenz. □

[27] Ein Maß ist bereits durch seine Werte auf einem durchschnittstabilen Erzeuger eindeutig festgelegt.

Satz 93. *$X_n \xrightarrow{P} X$ impliziert $X_n \Rightarrow X$. Ist X konstant, so gilt auch die Umkehrung.*

Beweis. Seien $X_n \xrightarrow{P} X$. Wir zeigen die Behauptung durch Widerspruch. Angenommen für $f \in C_b(E)$ gilt $\lim_{n\to\infty} E[f(X_n)] \neq E[f(X)]$. Dann gibt es eine Teilfolge n_k und ein $\varepsilon > 0$, so dass

$$\lim_{k\to\infty} |E[f(X_{n_k})] - E[f(x)]| > \varepsilon.$$

Da $X_n \xrightarrow{P} X$ gibt es eine Teilfolge (n_{k_ℓ}), $\ell = 1, 2, \ldots$, so dass $X_{n_{k_\ell}} \xrightarrow{f.s.} X$. Mit majorisierter Konvergenz folgt

$$\lim_{e\to\infty} E[f(X_{n_{k_\ell}})] = E[f(X)],$$

ein Widerspruch.

Ist umgekehrt $X = c \in Z$. Dann ist $x \mapsto d(x,c) \wedge 1 \in C_b(E)$ und somit

$$\mathbb{E}[d(X_n, c) \wedge 1] \to E[d(X, c) \wedge 1] = 0,$$

also $X_n \overset{P}{\to} X$ wegen der Definition von stochastischer Konvergenz.

□

Wir benötigen folgendes nützliche Resultat was man mit dem Fubini-Theorem beweist (!): Für reellwertiges $X \geq 0$ ist

$$E[X] = \int_0^\infty P(X > t)dt$$

Theorem 94 (Portmanteau). *Seien $X, X_1, X_2, \ldots$ Zufallsvarialbe mit Werten in E. Dann sind äuqivalent*

(i) $X_n \Rightarrow X$
(ii) $E[f(X_n)] \Rightarrow E[f(X)]$ *für alle Lipschitz-stetigen und beschränkten* f
(iii) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ *für alle offenen* $G \subseteq E$
(iv) $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ *für alle abgeschlossenen* $F \subseteq E$
(v) $\lim_{n\to\infty} P(X_n \in B) = P(X \in B)$ *für alle* $B \in \mathscr{B}(E)$ *mit* $P(X \in \partial B) = 0$.

Die $B \in \mathscr{B}(E)$ mit $P(X \in \partial B) = 0$ heißen auch $P_\# X$-randlos. $\partial B = \overline{B} \setminus B^o$ ($\overline{B}$ = Abschluss, B^o = Innere von B).

Beweis.

$(i) \Rightarrow (ii)$: klar

$(ii) \Rightarrow (iv)$: Sei $F \subseteq E$ abgeschlossen und (f_m) Lipschitz-stetige Funktion, so dass $f_m \downarrow \mathbb{1}_F$ (etwa $f_m(x) = (1 - m \cdot d(x,F))^+$). Dann folgt

$$\begin{aligned}\limsup P(X_n \in F) &\leq \inf_{k\geq 1} \limsup_{n\to\infty} E[f_k(X_n)]] \overset{(ii)}{=} \inf_{k\geq 1} E[f_k(X)] \\ &= P(X \in F)\end{aligned}$$

$(iv) \Leftrightarrow (iii)$: Setze $G = E \setminus F$ bzw. $F = E \setminus G$.

$(iii) \Rightarrow (i)$: Sei $f \geq 0$ stetig. Dann ist

$$\begin{aligned}E[f(X)] = \int_0^\infty P(f(X) > t)\,dt &\overset{(iii)}{\leq} \int_0^\infty \liminf P(X_n \in G) dt \\ &\overset{\text{Fatou}}{\leq} \liminf \int_0^\infty P(f(X_n) > t) dt = \liminf E[f(X_n)]\end{aligned} \tag{95}$$

Ist nun $-c < f < c$, so ist $c - f \geq 0$ und wir können das obige Resultat anwenden. Das heißt

$$\begin{aligned}\limsup E[f(X_n)] &= c - \liminf E[-f(X_n) + c] \leq c - E[-f(X) + c] \\ &= E[f(X)] \leq \liminf E[f(X_n)]\end{aligned}$$

(wobei wir für den letzten Schrift $f + c$ betrachtet haben) und somit folgt (i).

$(iv) \Rightarrow (v)$: Für $B \in \mathscr{B}(E)$:

$$P(X \in B^o) \overset{(iii)}{\leq} \liminf \underbrace{P(X_n \in B^o)}_{\leq P(X_n \in \overline{B})} \leq \limsup P(X_n \in \overline{B}) \leq P(X \in \overline{B}).$$

Ist $P(X \in \partial B) = 0$, so folgt $P(X \in B^o) = P(X \in \overline{B})$, also $P(X_n \in B) \to P(X \in B)$.

$(v) \Rightarrow (iv)$: Sei $F \subseteq E$ abgeschlossen. Wir setzen $F^\varepsilon := \{x \in E : d(x,F) \leq \varepsilon\}$ für $\varepsilon > 0$. Die Mengen $\partial F^\varepsilon \subseteq \{x : d(x,F) = \varepsilon\}$ sind disjunkt, also gilt

$$P(X \in \partial F^\varepsilon) = 0$$

für Lebesgue-fast alle ε. Wir wählen eine Folge $(\varepsilon_k) \downarrow 0$, so dass

$$\begin{aligned}\limsup P(X_n \in F) &\leq \inf_{k\geq 1} \limsup P(X_n \in F^{\varepsilon_k}) \\ &= \inf_{k \geq 1} P(X \in F^{\varepsilon_k}) = P(X \in F),\end{aligned} \tag{96}$$

wobei wir im letzten Schritt die σ-Stetigkeit von P ausgenutzt haben.

$\square$

Korollar 97. *Seien $X, X_1, X_2, \ldots$ reellwertige Zufallsvariable mit Verteilungsfunktion $F, F_1, F_2, \ldots$ Dann*

$$X_n \Rightarrow X \Leftrightarrow F_n(x) \to F(x) \text{ für alle Stetigkeitsstellen } x \text{ von } F.$$

Beweis. „$\Rightarrow$“ Ist x Stetigkeitsstelle von F, so ist zunächst $P(X \in \partial(-\infty, x]) = P(X = x) = 0$. Nach dem Portmanteau Theorem, Theorem 94(v), folgt aus $X_n \Rightarrow X$, dass $F_n(x) = P(X_n \in (-\infty, x]) \to P(X \in (-\infty, x]) = F(x)$.

„$\Leftarrow$“ Wir verwenden (ii) von Theorem 94: Ohne Einschränkung sei $|f| \leq 1$, f stetig und mit Lipschitz-Konstante 1. Für $\varepsilon > 0$ wählen wir $N \in \mathbb{N}$ und Stetigkeitspunkte $y_1, \ldots, y_N$ von F so, dass

$$F(y_0) < \varepsilon, \qquad F(y_N) > 1 - \varepsilon$$

und $y_i - y_{i-1} < \varepsilon$, $i = 1, \ldots, N$. Dann gilt nach Voraussetzung $F_n(y_i) \to F(y_i)$ und

$$f \leq \mathbb{1}_{(-\infty, y_0)} + \mathbb{1}_{(y_N, \infty)} + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon) \mathbb{1}_{(y_i, y_{i+1}]}.$$

Es folgt, dass

$$\begin{aligned}
&\limsup_{n\to\infty} E[f(X_n)] \\
&\leq \limsup \Big[F_n(y_0) + 1 - F_n(y_N) + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon)\big(F_n(y_{i+1}) - F_n(y_i)\big) \Big] \\
&\leq 3\varepsilon + \sum_{i=1}^{N-1} f(y_i)\big(F(y_{i+1}) - F(y_i)\big) \\
&\leq 4\varepsilon + E[f(x)].
\end{aligned}$$

Verwenden wir $1 - f$, so erhalten wir eine analoge Aussage für den $\liminf$ und die Behauptung folgt mit $\varepsilon \to 0$. □

Korollar 98 (Slutsky)**.**

$$X_n \Rightarrow X, \qquad d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \Rightarrow X.$$

Beweis. Übungsaufgabe □

Theorem 99 (Continuous mapping theorem). *Seien (E, d) und (E', d') metrische Räume, $f : E \to E'$ messbar und $U_f = \{x \in E : f \text{ nicht stetig an } X\}$ messbar.*

(i) Gilt $P_n \Rightarrow P$ und $P(U_f) = 0$, so folgt $f_* P_n \Rightarrow f_* P$
(ii) Gilt $X_n \Rightarrow X$ und $P(X \in U_f) = 0$, so folgt $f(X_n) \Rightarrow f(X)$

Beweis. (ii) folgt direkt aus (i) mit $P_n = X_{n*} P$

Für Teil (i) betrachte $G \subseteq E'$ offen und $x \in f^{-1}(G) \cap U_f^c$.

Da f stetig in x ist, ist $f^{-1}(G)$ offen und so existiert $\delta > 0$ dass $f(y) \in G$ für alle $y \in E$ mit $d(x, y) < \delta$.

Dann ist $f^{-1}(G) \cap U_f^c \subseteq (f^{-1}(G))^o$ und mit dem Portmanteau-Theorem 94 ($i \Rightarrow iii$) folgt

$$\begin{aligned} f_* P(G) = P(f^{-1}(G) &\overset{P(\mu_f)=0}{=} P(f^{-1}(G) \cap U_f^c) \leq P((f^{-1}(G))^0) \\ &\leq \liminf_{n \to \infty} P_n((f^{-1}(G))^0) \leq \liminf_{n \to \infty} P_n((f^{-1}(G))) \\ &= \liminf f_* P_n(G) \end{aligned}$$

und wieder mit dem Portmanteau-Theorem 94 ($iii \Rightarrow i$) folgt die Behauptung. □

Aufgabe 3. Ist Z separabel, so ist U_f messbar.

Als wichtiges Resultat erhalten wir den Satz von Skorokhod, der einen Zusammenhang zwischen schwacher und fast sicherer Konvergenz bildet. [28]

[28] Hinweis: Kallenberg verwendet das Theorem ohne die Vollständigkeit, damit ist die Existenz von iid Variablen etwas schwerer!

Theorem 100 (Skorokhod). *Sei (E, r) vollständig und separabel, $X, X_1, \ldots$ Zufallsvariable mit Werten in E. Dann sind äquivalent*

(i) $X_n \Rightarrow X$
(ii) Es gibt ein $(\Omega', \mathcal{A}', P')$ mit Zufallsvariablen Y, Y_1, Y_2 und $Y_n \xrightarrow{f.s.} Y$ sowie $Y \overset{\mathscr{L}}{=} X$, $Y_i \overset{\mathscr{L}}{=} X_i$, $i = 1, 2, \ldots$

Beweis. „$\Leftarrow$“ klar.

„$\Rightarrow$“ Zunächst nehmen wir $E = \{1, \ldots, m\}$ an. Setze $p_k = P(X = k)$ und $p_k^n = P(X_n = k)$, $k \in E$. Nimmt man $U \sim U[0, 1]$ unabhängig von X, so kann man leicht Zufallsvariablen $Y_n \overset{\mathscr{L}}{=} X_n$ konstruieren,so dass $Y_n = k$, falls $X = k$ und $U \leq \frac{p_k^n}{p_k}$:

Nehmen wir an, dass $p_k^n \uparrow p_k$ folgt $Y_n \Rightarrow X$, und da es auch nur endlich viele Werte sind sogar f. s.- sind allerdings die $p_k^n > p_k$, so müssen wir die Wahrscheinlichkeit größer machen, betrachten also

$$Y_n = k \qquad \text{falls } \big(X = k \text{ und } U \le \frac{p_k^n}{p_k}\big) \text{ oder } U' \in [0, (p_k^n - p_k)^+]$$

mit einer weiteren unabhängigen Zufallsvariablen $U' \sum U[0,1]$. Auch hier folgt wieder fast sichere Konvergenz.

Sei nun E allgemein. Wir fixieren $p \in \mathbb{N}$ und teilen E in Borel-Mengen $B_1, B_2, \dots$ mit Durchmesser kleiner 2^{-p} mit $P(X \in \partial B_n) = 0$.

Nun wählen wir m groß genug so, dass

$$P\Big(X = \bigcup_{k \le m} B_k\Big) < 2^{-p}$$

und setzen $B_0 := E \setminus \bigcup_{k \le m} B_k$.

Für $k = 0, \dots, m$ setzen wir $Y = k$ falls $X \in B_k$ und $Y_n = k$ falls $X_n \in B_k, n \in \mathbb{N}$. Dann gilt $Y_n \Rightarrow Y$ und wegen des Resultats über endliches E finden wir $Y_n^N \overset{\mathscr{L}}{=} Y_n$, so dass $Y_n^n \Rightarrow Y f.s.$ □

Lemma 101. *Wir betrachten* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. *Sind* $P_1, P_2, \dots \in \mathscr{P}(\mathbb{R})$ *und* $\mu \in \mathscr{P}_{\le 1}(\mathbb{R})$ *mit* $P_n \Rightarrow_v \mu$, *so gilt* $\mu(\mathbb{R}) \le 1$.

Beweis: Wir wählen $(f_n) \in C_c(\mathbb{R})$ mit $f_n \uparrow 1$. Mit monotoner Konvergenz folgt

$$\mu(\mathbb{R}) = \sup_{n \ge 1} \int f_n d\mu = \sup_{n \ge 1} \lim_{k \to \infty} \int f_n dP_k \le 1.$$ □

Dass $\mu(\mathbb{R}) < 1$ möglich ist hatten wir bereits in Beispiel 91 gesehen.

Schwache Konvergenz auf $\mathbb{R}$

In endlich-dimensionalen Räumen ist Kompaktheit ein tragfähiger Begriff. In unendlichdiemsionalen Räumen ist dieser Begriff deutlich subtiler. So ist etwa in vielen Topologien der Einheitsball nicht kompakt.

Neben dem Begriff der *Kompaktheit*[29] gibt es einen eigenständigen Begriff der *Folgenkompaktheit* (jede Folge hat eine konvergente Teilfolge).

Der folgende Satz zeigt zunächst zweierlei: $\mathscr{P}(\mathbb{R})$ ist nicht folgenkompakt: Es gibt zwar konvergente Teilfolgen, diese verlassen aber möglicherweise den Raum der Wahrscheinlichkeitsmaße. Zweitens: Es ist überhaput nicht zu erwarten, dass jede Folge von Wahrscheinlichkeitsmaßen eine schwach konvergente Teilfolge hat (wir erinnern uns an δ_n).

[29] In einem topologischen Raum heißt eine Menge kompakt, falls jede offene Überdeckung eine endliche Teilüberdeckung besitzt. Im $\mathbb{R}^n$ sagt der Satz von Heine-Borel, dass eine Menge genau dann kompakt ist, wenn sie abgeschlossen und beschränkt ist.

Durch den folgenden Satz erlangen wir immerhin Folgenkompaktheit in $\mathscr{P}_{\geq 1}(\mathbb{R})$ bezüglich der *vagen* Topologie.

Theorem 102 (Helly-Bray). *Seien $P_1, P_2, \dots \in \mathscr{P}(\mathbb{R})$. Dann gibt es eine Teilfolge $(P_{n_k})_{k\geq 1}$ und $\mu \in \mathscr{P}_{\leq 1}(\mathbb{R})$, so dass*

$$P_n \Rightarrow_v \mu.$$

Beweis. Die Werte der Verteilungsfunktion $F_1, F_2, \dots$ der (P_n) leben auf dem kompakten Intervall $[0,1]$. Wir nutzen eine Abzählung (x_n) von $\mathbb{Q}$ und finden wegen der Kompaktheit von $[0,1]$ eine Teilfolge (F_{n_k}), so dass

$$F_{n_k}(x_i) \underset{k\to\infty}{\longrightarrow} G(x_i) \qquad i \geq 1$$

mit $G(x_i) \in [0,1]$ für alle $i \geq 1$. Dies zeigen wir mit einem Diagonalfolgenargument[30]: Zunächst finden wir iterativ Indexmengen $I_1 \supset I_2 \supset \dots$

[30] Hier wählen wir ein Diagonalfolgenargument um aus abzählbaren vielen konvergenten Folgen eine konvergente Teilfolge zu konstruieren.

$$\begin{aligned}
I_1 &= \{n_1^1, n_2^1, \dots\} \subset \mathbb{N}, \text{ so dass } F_{n_k^1}(x_1) \text{ konvergiert mit } n_k^1 \to \infty, \\
I_2 &= \{n_1^2, n_2^2, \dots\} \subset \mathbb{N}, \text{ so dass } F_{n_k^2}(x_2) \text{ konvergiert mit } n_k^2 \to \infty, \\
&\vdots \qquad\qquad \vdots
\end{aligned}$$

Damit konvergiert $(F_{n_k^j}(x_i))_k$ für $j \leq i$, den Grenzwert bezeichnen wir mit $G(x_i)$. Der Diagonalfolgentrick ist, die Diagonalfolge $(n_k^k)_k$ zu betrachten. Für jedes $k \in \mathbb{N}$ existiert $j \geq k$, so dass $n_j^j \in I_j \subseteq I_k$, so dass $(F_{n_j^j}(x_k))_j \to G(x_k)$.

Nun definieren wir

$$F(x) = \inf\{G(r) : r \in [0,1] \cap \mathbb{Q} : r > x\},$$

und mit (F_{n_k}) ist auch G und somit F monoton sowie rechtsstetig. Damit existiert ein zugehöriges Maß μ, so dass $\mu((x,y]) = F(y) - F(x)$, $x < y \in \mathbb{R}$.

Wir zeigen noch $P_n \Rightarrow_{\mathrm{v}} \mu$. Dazu sei $f \in C_c(\mathbb{R})$. Ohne Einschränkung sei $f \geq 0$. Nach Konstruktion gilt $F_n(x) \to F(x)$ an allen Stetigkeitsstellen von F. Da F wachsend und rechtsstetig ist, gibt es nur abzählbar viele Unstetigkeitsstellen $D \subseteq \mathbb{R}$. Es folgt für jedes $U = [x,y]$ mit $x, y \notin D$

$$P_n(U) \to \mu(U)$$

ebenso wie für endliche Vereinigungen, welche wir mit $\mathcal{V}$ bezeichnen.

Wir wählen $B \subseteq \mathbb{R}$ offen und $(U_k), (V_k) \subseteq \mathcal{V}$, so dass $U_k \uparrow B$, $V_k \downarrow \overline{B}$. Dann ist

$$\begin{aligned} \int_B d\mu &= \lim_{k\to\infty} \mu(U_k) = \lim_{k\to\infty} \liminf_{n\to\infty} P_n(U_k) \leq \liminf_{n\to\infty} P_n(B) \\ &\leq \limsup_{n\to\infty} P_n(B) \leq \lim_{k\to\infty} \limsup_{n\to\infty} P_n(V_k) = \lim_{k\to\infty} \mu(V_k) = \int_{\overline{B}} d\mu \end{aligned}$$

Da $\mu(f = t) = 0$ für Lebesgue-fast alle t, folgt mit majorisierter Konvergenz

$$\begin{aligned} \int f d\mu = \int_0^\infty \mu(f > t) dt &\leq \int_0^\infty \liminf P_n(f > t) dt \\ &\leq \liminf \int_0^\infty P_n(f > t) dt = \liminf \int f dP_n \\ &\leq \limsup \int f dP_n = \limsup \int_0^\infty P_n(f > t) dt \\ &\leq \int \limsup P_n(f > t) dt \leq \int_0^\infty \mu(f \geq t) dt \\ &= \int f d\mu, \end{aligned}$$

also $\lim \int f dP_n = \int f d\mu$. □

Das folgende Korollar zeigt, dass bereits die Eindeutigkeit des Häufungspunktes, mit der Eigenschaft dass es sich um ein Wahrscheinlichkeitsmaß handelt, schwache Konvergenz impliziert.

Korollar 103. *Seien $P_1, P_2, \dots \in \mathscr{P}(\mathbb{R})$ und sei $P \in \mathscr{P}(\mathbb{R})$ der einzige Häufungspunkt von (P_n) bezüglich der schwachen Konvergenz. Dann gilt*

$$P_n \Rightarrow P.$$

Beweis. Angenommen, die Folge (P_n) konvergiere nicht. Wir bezeichnen mit F_n und F die zu P_n und P gehörigen Verteilungsfunktionen.

Dann gibt es einen Stetigkeitspunkt x von F, so dass $F_n(x) \not\to F(x)$. Damit gibt es also eine Teilfolge (n_k) und $a \neq F(x)$, so dass $F_{n_k} \to a$.

Nach Theorem 102 von Helly-Bray gibt es eine Teilfolge $(n_{k_l})_k$ welche vage konvergiert. Da P der einzige Häufungspunkt von (P_n) ist, folgt $F_{n_{k_l}}(x) \to F(x)$, ein Widerspruch. □

Straffheit und relative Kompaktheit

In diesem Abschnitt beschäftigen wir uns mit der Fragen wann eine Folge von Wahrscheinlichkeitsmaßen einen schwachen Grenzwert besitzt, oder wenigstens einen Häufungspunkt. Dies muss nicht automatisch der Fall sein, wie das Beispiel (δ_n) zeigt. Damit ein Häufungspunkt existiert müssen wir sicherstellen, dass keine Masse ins unendlich abwandert - was wir durch den Begriff der Straffheit erreichen.

Auf einem topologischen Raum $(\Omega, \mathcal{O})$ heißt $K \subseteq \Omega$ *kompakt*, falls jede offene Überdeckung von K eine endliche Teilüberdeckung besitzt,

Nun betrachten wir wieder einen allgemeinen metrischen Raum (E, d). Mit $\mathcal{K}$ bezeichnen wir die kompakten Mengen in E.

Definition 104. Eine Familie $(P_i)_{i\in I} \subseteq \mathscr{P}(E)$ heißt *straff*, falls

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} P_i(K) = 1.$$

Eine Familie $(X_i)_{i\in I}$ von Zufallsvarialben heißt *straff*, falls $(X_{i*}P)_{i\in I}$ straff sind, also

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} P(X_i \in K) = 1.$$

Bemerkung 105. Äquivalente Formulierungen:

(i) Für alle $\varepsilon > 0$ gibt es $K_\varepsilon \in \mathcal{K}$, so dass $P_i(K^c) < \varepsilon$ für alle $i \in I$.

(ii) Ist $E = \mathbb{R}^d$, so ist $(P_i)_{i\in I}$ straff $\Leftrightarrow$

$$\sup_{r>0} \inf_{i\in I} P_i(B_r(0)) = 1 \qquad (B_r(0) = \{x : \|x\| \leq r\})$$

(iii) Ein einzelnes Maß P ist straff, falls (E, d) vollständig und separabel ist. Damit folgt dies für jede endliche Familie von Wahrscheinlichkeitsmaßen. Das kann man mit Hilfe Lemma 114 zeigen.

Auf einem polnischen Raum kann man die volle Wahrscheinlichkeit durch kompakte Mengen ausschöpfen, was sich als sehr nützlich erweisen wird.[31] Der Beweis hiervon folgt in Lemma 114.

[31] Ein W-Maß heißt von **innen regulär**, wenn man die volle Masse durch kompakte Mengen approximieren kann. Das ist auf einem polnischen Raum möglich.

Lemma 106. *Eine abzählbare Familie $(P_n)_{n\geq 1} \subseteq \mathscr{P}(E)$ auf einem polnischen Raum E ist genau dann straff, wenn*

$$\sup_{K\in\mathcal{K}} \liminf_{n\geq 1} P_n(K) = 1. \tag{107}$$

Beweis. „$\Rightarrow$“: klar, da $\liminf P_n(K) \geq \inf P_n(K)$

„$\Leftarrow$“: Sei $\varepsilon > 0$ und K so, dass $\liminf P_n(K) \geq 1 - \frac{\varepsilon}{2}$.

Wir wählen m so, dass $\inf_{n \geq m} P_n(K) \geq 1 - \varepsilon$ und $K_1, \ldots, K_m \in \mathcal{K}$ so, dass $P_n(K_n) \geq 1 - \varepsilon$, $n = 1, \ldots, m$. Da $K' = K \cup K_1 \cup \cdots \cup K_m \in \mathcal{K}$ wieder kompakt ist und $\inf P_n(K') \geq 1 - \varepsilon$ folgt die Straffheit. (Zu jedem K finde ich $K' \geq K$, so dass $\inf P_n(K') \geq 1 - \varepsilon$, siehe Bemerkung 105 (i)) □

Beispiel 108. (i) Ist E kompakt, so ist jede Familie von Wahrscheinlichkeitsmaßen straff.

(ii) Sind $(X_i)_{i \in I}$ reellwertig mit

$$\sup_{i \in I} E[|X_i|] < \infty,$$

so ist (X_i) straff: Denn es gilt nach der Markov-Ungleichung

$$\inf_{r>0} \sup_{i \in I} P(|X_i| \geq r) \leq \inf_{r>0} \sup_{i \in I} \frac{E[|X_i|]}{r} = 0$$

(iii) $(\delta_n)_{n \geq 1}$ ist nicht straff.

Das folgende Lemma zeigt nun, dass auf den reellen Zahlen Straffheit genau der richtige Begriff ist, um aus vager Konvergenz auf schwache Konvergenz zu schließen.

Lemma 109. *Seien $P_1, P_2, \ldots \in \mathscr{P}(\mathbb{R})$ und $\mu \in \mathscr{P}_{\leq 1}(\mathbb{R})$ mit*

$$P_n \Rightarrow_v \mu.$$

Dann ist $\mu(\mathbb{R}) = 1$ genau dann, wenn (P_n) straff ist. In diesem Fall folgt $P_n \Rightarrow \mu$.

Beweis: Sei $B_r = \{x \in \mathbb{R} : |x| \leq r\}$ der Ball um die Null mit dem Radius r. Für jedes $r > 0$ wählen wir $g_r \in C_c(\mathbb{R})$ mit $\mathbb{1}_{B_r} \leq g_r \leq \mathbb{1}_{B_{r+1}}$. Dann ist (P_n) genau dann straff, wenn

$$\sup_{r>0} \liminf_{n \to \infty} \int g_r dP_n = 1. \tag{110}$$

„$\Rightarrow$“: Mit Stetigkeit von unten (von μ) ist

$$1 = \sup_{r>0} \mu(B_r) = \sup_{r>0} \int g_r d\mu = \sup_{r>0} \liminf_{n \to \infty} \int g_r dP_n,$$

also folgt (110) und somit Straffheit.

„$\Leftarrow$“: Mit Lemma 106 (im letzten Schritt) folgt

$$\mu(\mathbb{R}) = \sup_{r>0} \mu(B_r) = \sup_{r>0} \liminf_{n\to\infty} \int g_r dP_n = 1.$$

Es bleibt, die schwache Konvergenz zu zeigen: Sei also (P_n) straff und $f \in C_b(\mathbb{R})$. Dann ist

$$\begin{aligned}
&\limsup_{n\to\infty} \Big| \int f_n dP_n - \int f d\mu \Big| \\
&\quad \le \inf_{r>0} \limsup_{n\to\infty} \Big(\Big| \int (f - f g_r) dP_n \Big| + \Big| \int f g_r dP_n - \int f g_r d\mu \Big| + \int (f - f g_r) d\mu \Big) \\
&\quad \le \|f\|_\infty \Big(\inf_{r>0} \limsup_{n\to\infty} P_n(B_r^c) + 0 + \inf_{r>0} \mu(B_r^c) \Big) = 0.
\end{aligned}$$

□

Korollar 111. *Seien $P, P_1, P_2, \ldots \in \mathscr{P}(\mathbb{R})$. Gilt $P_n \Rightarrow P$, so ist (P_n) straff.*

Kompaktheit und verwandte Begriffe

Nun sind wir daran interessiert, Straffheit auch auf algemeineren Räumen als $\mathbb{R}$ genau zu charakterisieren. Dazu werden wir in die topologische Trickkiste greifen müssen und bereiten zunächst einige Begriffe vor.

Definition 112. Auf einem topologischen Raum $(\Omega, \mathcal{O})$ heißt $K \subseteq \Omega$

(i) *kompakt*, falls jede offene Überdeckung von K eine endliche Teilüberdeckung besitzt,
(ii) *relativ kompakt*, falls $\overline{K}$ kompakt ist,
(iii) *relativ folgenkompakt*, falls es für jede Folge $\omega_1, \omega_2, \ldots \in K$ eine konvergente Teilfolge gibt.
(iv) Ist (Ω, d) ein metrischer Raum, so heißt K *total beschränkt*, falls es für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ und $\omega_1, \ldots, \omega_N \in K$ gibt, so dass $K \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$.

Nach dem Satz von Helly-Bray ist eine Familie von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ also immer relativ kompakt.

Satz 113. *Sei (E,d) ein metrischer Raum und $\mathcal{O}$ die von d erzeugte Topologie.*

(i) K ist relativ kompakt
(ii) Sind $F_i \subseteq \overline{K}$ abgeschlossen, $i \in I$, $\bigcap_{i\in I} F_i = \varnothing$, dann gibt es endliches $J \subseteq I$ mit $\bigcap_{i\in J} F_i = \varnothing$.
(iii) K ist relativ folgenkompakt.
(iv) K ist total beschränkt.

Dann gilt (iv)$\Leftarrow$ (i)$\Leftrightarrow$ (ii) $\Rightarrow$ (iii).

Ist (E,d) separabel, so folgt (iii) $\Rightarrow$ (ii).
Ist (E,d) vollständig, so folgt (iv) $\Rightarrow$ (iii).

Ist der Raum (E,d) polnisch, so sind die Eigenschaften äquivalent.

Beweis. *(i) $\Rightarrow$ (iv):* Sei $\overline{K}$ kompakt und $\varepsilon > 0$. Wir wählen die offene Überdeckung $\bigcup_{\omega\in K} B_\varepsilon(\omega)$ von $\overline{K}$. Dann gibt es eine endliche Teilüberdeckung, also gibt es $\omega_1,\dots,\omega_N$, so dass

$$\overline{K} \subseteq \bigcup_{n=1}^{N} B_\varepsilon(\omega_n).$$

Da ε beliebig war, folgt, dass K total beschränkt ist.

(i) $\Rightarrow$ (ii): Seien $F_i \subseteq \overline{K}$ abgeschlossen und $\bigcap_{i\in I} F_i = \varnothing$. Dann ist $\bigcup F_i^c = (\bigcap F_i)^c = \Omega$. Dies ist natürlich eine offene Überdeckung von $\overline{K}$, also gibt es endliches $J \subseteq I$, so dass $\overline{K} \subseteq \bigcup_{j\in J} F_j^c$.

Außerdem ist dann $\left(\bigcup_{j\in J} F_j^c\right)^c = \bigcap_{j\in J} F_j \subseteq \overline{K}^c$, aber $F_i \subseteq \overline{K}$, also muss $\bigcap_{j\in J} F_j = \varnothing$ sein.

(ii) $\Rightarrow$ (i): Seien O_i offen und $\overline{K} \subseteq \bigcup_{i\in I} O_i$. Wir setzen $F_i = O_i^c \cap \overline{K}$, so dass $F_i^c \in \mathcal{O}$ und $\bigcap_{i\in I} F_i = \overline{K} \cap (\bigcup O_i)^c = \varnothing$. Also gibt es nach Voraussetzung (ii) endliches $J \subseteq I$ mit $\bigcap_{j\in J} F_j = \varnothing$. Dann ist

$$\overline{K}^c \cup \bigcup_{j\in J} O_i = \bigcup_{j\in J} F_j^c = \Omega,$$

also $\overline{K} \subseteq \bigcup_{j\in J} O_j$ und somit $\overline{K}$ kompakt.

(ii) $\Rightarrow$ (iii): Seien $\omega_1, \omega_2, \dots \in K$. Wir setzen $F_n = \overline{\{\omega_n, \omega_{n-1}, \dots\}} \subseteq \overline{K}$. Angenommen es gibt keine konvergente Teilfolge von (ω_n). Dann ist $\bigcap_{n=1}^\infty F_n = \varnothing$.

Mit (ii) folgt dann, dass es ein $N \in \mathbb{N}$ gibt mit

$$\varnothing = \bigcap_{n=1}^{N} F_n = F_N,$$

ein Widerspruch zu $F_N \neq \emptyset$.

(iii) ⇒ (ii): (falls (E,d) separabel ist) Diesen Teil unterteilen wir in zwei Schritte. Da der Raum separabel ist gibt es abzählbares E' mit $\overline{E'} = E$ und $\mathscr{B} = \{B_{\frac{1}{n}}(\omega) : \omega \in E', n \in \mathbb{N}\}$. Dann ist $\mathscr{B}$ abzählbare Basis von $\mathcal{O}$, und wir schreiben $\mathscr{B} = \{C_1, C_2, \dots\}$.

Nun kommen wir zum zweiten Schritt. Angenommen $\overline{K}$ ist nicht kompakt: Dann gibt es $(A_i)_{i\in I} \subseteq \mathscr{B}$ mit $\overline{K} \subseteq \bigcup_{i\in I} A_i$ und es gibt keine endliche Teilüberdeckung. Wir setzen für $i \in I$

$$J_i = \{j \in \mathbb{N} : C_j \subseteq A_i\} \subseteq \mathbb{N}$$

sowie $J := \bigcup_{i\in I} J_i \subseteq \mathbb{N}$. Dann ist $A_i = \bigcup_{j\in J_i} C_j$, also

$$K \subseteq \bigcup_{i\in I} A_i = \bigcup_{i\in I}\bigcup_{j\in J_i} C_j = \bigcup_{j\in J} C_j.$$

Damit sind die (C_j) eine offene Überdeckung. Da es für (A_i) keine Teilüberdeckung gibt, gibt es ebenso keine für die (C_i). Wir wählen $\omega_n \in \overline{K} \setminus \bigcup_{j\in J, j\leq n} C_j$. Da K relativ folgenkompakt ist, hat (ω_n) einen Häufungspunkt in $\omega \in \overline{K}$. Da $\overline{K} \underset{j\in J}{\subseteq} C_j$, gibt es $k \in J$ mit $\omega \in C_k$. Da C_k offen liegen unendlich viele der w_n in C_k, andererseits ist $\omega_i \notin C_k$, $i \geq k$. Dies ist ein Widerspruch.

(iv) ⇒ (iii): (falls (E,d) vollständig) Seien $\omega_1, \omega_2, \dots \in K$. Wir konstruieren eine Teilfolge, die eine Cauchyfolge ist. Da E vollständig ist, ist die Folge auch konvergent und somit K relativ folgenkompakt.

Wir wählen $\varepsilon_n \downarrow 0$. Da K total beschränkt ist, gibt es $N \in \mathbb{N}$ ε_1-Bälle, die K überdecken. Mindestens einer dieser Bälle enthält unendlich viele der (ω_i). Diese haben dann höchstens den Abstand $2\varepsilon_1$. Wir wählen ein ω_{k_1} davon. Dieser ε_1-Ball wird durch endlich viele ε_2-Bälle überdeckt. Mindestens einer dieser Bälle enthält unendlich viele der (ω_i) und wir wählen $\omega_{k_2} \neq \omega_{k_1}$. Iterieren liefert (ω_{k_n}) mit $d(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m\wedge n}$ und somit die gesuchte Cauchyfolge.

□

Lemma 114. *Ist $(\Omega, \mathcal{O})$ polnisch und μ ein endliches Maß auf $\mathscr{B}(\mathcal{O})$, so existiert zu jedem $\varepsilon > 0$ eine kompakte Menge K, so dass $\mu(\Omega \setminus K) < \varepsilon$.*

Beweis: Sei $\varepsilon > 0$. Aus der Separabilität folgt

$$\Omega = \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(\omega_k^n) \quad \text{für alle } n \geq 1 \text{ und geeignete } (\omega_k^n)_{k\geq 1}.$$

Mit Stetigkeit von oben erhält man

$$0 = \mu\Big(\Omega \setminus \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(\omega_k^n)\Big) = \lim_{N\to\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^{N} B_{\frac{1}{n}}(\omega_k^n)\Big).$$

Demnach gibt es $N_n \in \mathbb{N}$ so dass $\mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)\Big) \leq \frac{\varepsilon}{2^n}$.

Die Menge $A := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)$ ist total beschränkt, also ist $\overline{A}$ kompakt nach Satz 113. Es folgt

$$\begin{aligned} \mu(\Omega \setminus \overline{A}) \leq \mu((\Omega \setminus A) &= \mu\Big(\bigcup_{k=1}^{\infty}\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)\Big)\Big) \\ &\leq \sum_{n=1}^{\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(\omega_k^n)\Big) \leq \varepsilon. \end{aligned}$$

□

Der Satz von Prohorov

Wir erinnern an einige Tatsachen:

$$K \text{ ist relativ kompakt } \Rightarrow K \text{ ist relativ folgenkompakt}$$

(Portmanteau)

$$(X_n) \text{ straff } \Rightarrow \liminf P(X_n \in G) \geq P(X \in G), \quad G \text{ offen}$$

Der *Satz von Prohorov* liefert eine wichtige Konsequenz aus Straffheit: Folgenkompaktheit (jede Folge hat eine konvergente Teilfolge). Für ein Familie von Maßen $(\mu_i)_{i\in I}$ ist relativ Folgenkompaktheit äquivalent zur Straffheit.

Theorem 115. *Sei (E, d) ein polnischer Raum. Dann gilt:*

$$(P_i)_{i\in I} \text{ ist straff} \Leftrightarrow (P_i)_{i\in I} \text{ ist relativ kompakt bzgl. der schwachen Topologie.}$$

Mit dem Satz von Prohorov wird eine nützliche Brücke zur Analysis geschlagen: Zum Beispiel charakterisiert der Satz von *Arzela-Ascoli* (welchen wir bald kennenlernen werden) relativ kompakte Mengen auf dem Raum $C(X)$ (stetige Funktionen $f : X \to \mathbb{R}$ versehen mit der Supremumsnorm), falls X ein kompakter topologischer Raum ist: $F \subset C(X)$ ist genau dann relativ kompakt wenn F gleichgradig stetig und punktweise beschränkt ist.

Bemerkung 116 (Relative Kompaktheit bezgl. der schwachen Topologie)**.** Im Folgenden Beweis zeigen wir relative Folgenkompaktheit. Dies ist aber ein einem polnischen Raum äquivalent zu relativer Kompaktheit (siehe etwa [32], Satz 8.4.13 oder ausführlicher in [33], Section 5).

[32] Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, 1977

[33] Patrick Billingsley. *Probability and Measure*. Wiley, 3 edition, 1995

Hierzu betrachten wir die *Prohorov-Metrik*. Sei (E,d) ein metrischer Raum und $\mathscr{B}(E)$ die zugehörige Borelsche σ-Algebra. Für eine Menge $B \in \mathscr{B}(E)$ betrachten wir ihre ε-Umgebung $B_\varepsilon = \{x \in E : d(x,B) < \varepsilon\}$. Für zwei Maße μ und ν definieren wir

$$d^*(\mu,\nu) := \inf\{\varepsilon > 0 : \text{ für alle } B \in \mathscr{B}(E) \text{ ist } \mu(B) \le \nu(B_\varepsilon) + \varepsilon\}.$$

Die Prohorov-Metrik ist nun definiert durch

$$d_P(\mu,\nu) := \max\{d^*(\mu,\nu), d^*(\nu,\mu)\}. \tag{117}$$

Diese Metrik bestimmt schwache Konvergenz: $d_P(\mu_n,\mu) \to 0$ impliziert $\mu_n \Rightarrow \mu$. Ist (E,d) separabel, so gilt auch die Umkehrung.

Außerdem gilt, dass (E,d) genau dann polnisch ist wenn der Raum der endlichen Maße $(\mathscr{M}(E), d_P)$, versehen mit der Prohorov-Metrik, polnisch ist.

Beweis. „$\Leftarrow$" Sei $(x_n)_{n\in\mathbb{N}}$ eine dichte Teilmenge in E. Wir zeigen zunächst: Aus relativer Folgenkompaktheit folgt $\forall \varepsilon > 0\ \exists N \in \mathbb{N}$ s. d.[34]

$$\inf_{i\in I} P_i\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big) > 1-\varepsilon. \tag{118}$$

Angenommen, das gilt nicht. Dann existiert $\varepsilon > 0$ so dass für alle $x_1,\dots,x_N$, $N \in \mathbb{N}$ ein P_{i_N} existiert mit

$$P_{i_N}\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big) \le 1-\varepsilon.$$

Da nach Voraussetzung (P_i) relativ folgenkompakt ist, also gibt es eine Teilfolge $(P_{i_{N_M}})_M$, die schwach gegen ein $P \in \mathscr{P}(E)$ konvergiert. Wegen des Portmanteau-Theorems gilt

$$\begin{aligned}1 = P(E) &\underset{\text{Separabilität}}{=} \lim_{N\to\infty} P\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big)\\ &\le \lim_{N\to\infty} \liminf_{M\to\infty} P_{i_{N_M}}\Big(\bigcup_{k=1}^N B_\varepsilon(x_k)\Big) \le 1-\varepsilon\end{aligned}$$

und dies ist ein Widerspruch.

Nun zeigen wir, dass aus (118) Straffheit folgt: Sei $\varepsilon > 0$. Für $j \in \mathbb{N}$ wählen wir $x_1^j,\dots,x_{n_j}^j$ so, dass

$$\inf_{i\in I} P_i\Big(\bigcup_{k=1}^{n_j} B_{\frac{\varepsilon}{2^j}}(x_k)\Big) > 1-\frac{\varepsilon}{2^j}$$

Wir setzen

$$K := \bigcap_{j=1}^\infty \bigcup_{k=1}^{n_j} B_{\frac{\varepsilon}{2^j}}(x_k).$$

[34] $B_\varepsilon(x)$ ist offener Ball mit Radius ε um x

Dann ist K total beschränkt. Da E ein vollständiger metrischer Raum ist, ist K relativ kompakt, also ist $\overline{K}$ kompakt.

Außerdem ist

$$\sup_{i\in I} P_i(\overline{K}^c) \leq \sup_{i\in I} \sum_{j=1}^{\infty} P_i\Big(\bigcap_{k=1}^{n_j} B_{\frac{\varepsilon}{2^j}}(x_k)^c\Big) \leq \varepsilon \sum_{j=1}^{\infty} \frac{1}{2^j} = \varepsilon,$$

also ist (P_i) straff.

Nun betrachten wir die Umkehrung.

„$\Rightarrow$ “: Ziel ist es, zu zeigen, dass (P_i) relativ folgenkompakt ist: Sei also (P_k) eine Folge aus $(P_i)_{i\in I}$. Da (P_i) straff ist, können wir kompakte Mengen $K_1 \subseteq K_2 \subseteq \cdots$ so wählen dass

$$\inf_{n\geq k} P_n(K_j) \geq 1 - \frac{1}{j}.$$

Da (E,d) separabel ist, gibt es abzählbar viele $x_1, x_2, \ldots$ und $\varepsilon_1, \varepsilon_2, \ldots$, so dass $\{(B_{\varepsilon_k}(x_k))_{k\geq 1} : \varepsilon_k \in \mathbb{Q}, k \in \mathbb{N}\}$ $\mathscr{B}(E)$ erzeugen. Wir setzen

$$\mathcal{K} := \Big\{ \bigcup_{k=1}^{N} K_j \cap \overline{B_{\varepsilon_k}(x_k)} : N, j \in \mathbb{N}, 0 < \varepsilon_k \in \mathbb{Q}\Big\}.$$

Dann ist $\mathcal{K}$ ein Halbring und ein durchschnittstabiler Erzeuger von $\mathscr{B}(E) \cap (\bigcup_{j=1}^{\infty} K_j)$. Da $\mathcal{K}$ abzählbar ist, finden wir eine Teilfolge (P_{n_k}), so dass $(P_{n_k}(A))$ für alle $A \in \mathcal{K}$ konvergiert (verwende $[0,1]$ kompakt und Diagonlafolge).

Wir setzen $\mu(A) := \lim_{k\to\infty} P_{n_k}(A)$, $A \in \mathcal{K}$. Diesen Inhalt kann man eindeutig zu einem Maß μ auf $\mathscr{B}(E)$ fortsetzen. Es folgt:

$$1 \geq \mu(E) = \mu(\bigcup_{j=1}^{\infty} K_j) = \lim_{m\to\infty} \lim_{k\to\infty} P_{n_k}(\bigcup_{j=1}^{m} K_j) \geq \lim_{m\to\infty} \Big(1 - \frac{1}{m}\Big) = 1.$$

Für jedes offene A gilt

$$\begin{aligned}\mu(A) = \inf_{A_n\in\mathcal{K}, \bigcup A_n \supset A} \sum \mu(A_n) &= \inf_{A_n\in\mathcal{K}, \bigcup A_n \supset A} \lim_{k\to\infty} \sum P_{n_k}(A_n) \\ &\leq \liminf_{k\to\infty} P_{n_k}(A),\end{aligned}$$

und so folgt $P_{n_k} \Rightarrow P$ aus dem Portmanteau-Theorem. □

Der Zentrale Grenzwertsatz

In diesem Abschnitt beweisen wir den zentralen Grenzwertsatz von Lindeberg und Feller. Das wesentliche Hilfsmittel werden die charakteristischen Funktionen sein, die wir zunächst einführen. Anschließend beweisen wir den Satz von Lévy, welcher schwache Konvergenz mit Hilfe von charakteristischen Funktionen (oder allgemeiner: separierenden Funktionenklassen) verknüpft.

Charakteristische Funktionen

Definition 119. Eine Menge $\mathcal{M} \subseteq C(E)$ von stetigen Funktionen auf E heißt *punktetrennend*, falls für alle $x \neq y \in E$ ein $f \in \mathcal{M}$ existiert, so dass $f(x) \neq f(y)$.

$\mathcal{M}$ heißt *separierend*, falls aus $P, Q \in \mathscr{M}(E)$ und $E_P[f] = E_Q[f]$ für alle $f \in \mathcal{M}$ folgt, dass $P = Q$.

Beispiel 120. $C_b(E)$ ist punktetrennend und separierend: Wähle $f(x) = d(x,y) \wedge 1$. Ist $P \neq Q$, so gibt es einen offenen Ball A mit $P(A) \neq Q(A)$. Wir wählen $f_n \uparrow \mathbb{1}_A$ und erhalten $P(A) = \lim E_P[f_n]$ und $Q(A) = \lim E_Q[f_n]$. Wäre $C_b(E)$ nicht separierend, so erhielten wir $Q(A) = P(A)$, ein Widerspruch.

Wir nennen ein Mengensystem $\mathcal{M} \subset \mathcal{C}(E)$ **Algebra**, falls

(i)	$1 \in \mathcal{M}$	
(ii)	$\alpha f + \beta g \in \mathcal{M}$	für $\alpha, \beta \in \mathbb{R}$ und $f, g \in \mathcal{M}$
(iii)	$fg \in \mathcal{M}$	$f, g \in \mathcal{M}$

Theorem 121 (Stone-Weierstraß). *Sei (E,d) kompakt und $\mathcal{M} \subseteq C_b(E)$ eine punktetrennende Algebra. Dann liegt $\mathcal{M}$ dicht in $C_b(E)$ bzgl. der Supremumsnorm.*

Theorem 122. *Sei (E,d) vollständig und separabel. Ist $\mathcal{M} \subseteq C_b(E)$ punktetrennend und folgt aus $f,g \in \mathcal{M}$ dass $fg \in \mathcal{M}$, dann ist M separierend.*

Wir erhalten direkt

Satz 123. *Ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$ wird eindeutig durch die charakteristische Funktion bestimmt.*

Der Grund hierfür ist, dass , e^{iux}, $u \in \mathbb{R}^d$ eine separierende Familie auf dem $\mathbb{R}^d$ ist, was aus Theorem 122 folgt.

Beweis zu Theorem 122: Ohne Einschränkung ist $\mathcal{M}$ eine Algebra, denn aus $E_Q[f] = E_P[f]$ und ebenso für g folgt $\alpha E_Q[f] + \beta E_Q[f] = \alpha E_P[f] + \beta E_P[f]$, sowie $E_P[1] = E_Q[1]$.

Nach Lemma 114 können wir K kompakt wählen, so dass $P(K) > 1 - \varepsilon$ und $Q(K) > 1 - \varepsilon$. Nach dem Stone-Weierstraß-Theorem gibt es für $g \in C_b(E) \quad (g_n) \subseteq \mathcal{M}$ so dass

$$\sup_{x \in K} |g_n(x) - g(x)| \underset{n\to\infty}{\to} 0. \tag{124}$$

Wir betrachten $ge^{-\varepsilon g^2}$, so dass außerhalb von K diese Funktionenklasse schnell verschwindet. Durch Ableiten und Null setzen errechnet man das folgende Maximum

$$\sup_{x\in\mathbb{R}} xe^{-\varepsilon x^2} = e^{-1/2}\sqrt{\frac{1}{2\varepsilon}}.$$

Wir erhalten, dass

$$|E[ge^{-\varepsilon g^2}] - E[ge^{-\varepsilon g^2} 1_K]| \leq C\varepsilon^{-1/2} \cdot P(K^c) \leq C \cdot \sqrt{\varepsilon} \underset{\varepsilon\to 0}{\to} 0. \tag{125}$$

Wir schätzen ab:

$$\begin{aligned}
& \big|E_P[ge^{-\varepsilon g^2}] - E_Q[ge^{-\varepsilon g^2}]\big| \\
(1) \quad & \leq \big|E_P[ge^{-\varepsilon g^2}] - E_P[ge^{-\varepsilon g^2}\mathbb{1}_K]\big| \\
(2) \quad & \quad + \big|E_P[ge^{-\varepsilon g^2}\mathbb{1}_K] - E_P[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K]\big| \\
(3) \quad & \quad + \big|E_P[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K] - E_P[g_n e^{-\varepsilon g_n^2}]\big| \\
(4) \quad & \quad + \big|E_P[g_n e^{-\varepsilon g_n^2}] - E_Q[g_n e^{-\varepsilon g_n^2}]\big| \\
(5) \quad & \quad + \big|E_Q[g_n e^{-\varepsilon g_n^2}] - E_Q[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K]\big| \\
(6) \quad & \quad + \big|E_Q[g_n e^{-\varepsilon g_n^2}\mathbb{1}_K] - E_Q[ge^{-\varepsilon g^2}\mathbb{1}_K]\big| \\
(7) \quad & \quad + \big|E_Q[ge^{-\varepsilon g^2}\mathbb{1}_K] - E_Q[ge^{-\varepsilon g^2}]\big|
\end{aligned}$$

Dann gilt (1), (3), (5), (7) $\to 0$ wegen (125) und (2) und (6) $\to 0$ wegen (124). Es folgt

$$|E_P[g] - E_Q[g]| = \lim_{\varepsilon\to 0} \left| E_P[ge^{-\varepsilon g^2}] - E_Q[ge^{-\varepsilon g^2}] \right| = 0.$$

Da $g \in \mathcal{C}_b(E)$ beliebig und $C_b(E)$ separierend folgt $P = Q$. □

Bemerkung 126. Wir erhalten unmittelbar

(i) Auf dem $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ bestimmen charakteristische Funktionen eindeutig das Wahrscheinlichkeitsmaß.
(ii) Eine Familie reellwertiger Zufallsvariablen $(X_i)_{i\in J}$ ist genau dann unabhängig, wenn für alle endlichen $J \subseteq I$

$$E\Big[\prod_{j\in J} e^{iu_j X_j}\Big] = \prod_{j\in J} E[e^{iu_j X_j}], \quad (u_j)_{j\in J} \subseteq \mathbb{R}^0.$$

Aus Satz 123 erhält man unmittelbar das Cramér-Wold Device.

Korollar 127. *Seien X, Y d-dimensionale Zuvallsvariablen. Dann gilt*

$$X \stackrel{\mathscr{L}}{=} Y \qquad \Longleftrightarrow \qquad \langle t, X\rangle \stackrel{\mathscr{L}}{=} \langle t, Y\rangle \qquad \forall t \in \{t \in \mathbb{R}^d : \| t \| = 1\}.$$

Der Satz von Lévy

Als nächsten großen Schritt beweisen wir den Satz von Lévy . Der Satz von Lévy wird schwache Konvergenz vollständig mit Hilfe von charakteristischen Funktionen klassifizieren.

Satz 128. *Sei (E, d) vollständig und separabel und $P, P_1, P_2, \ldots \in \mathscr{M}(E)$.*

Dann sind äquivalent:

(i) $P_n \Rightarrow P$
(ii) $(P_n)_{n\geq 1}$ *ist straff und es gibt eine separierende Familie* $\mathcal{M} \subseteq C_b(E)$, *s. d.*

$$E_{P_n}[f] \to E_P[f] \quad \forall f \in \mathcal{M}.$$

Beweis. „$\Rightarrow$“ klar

„$\Leftarrow$“ Angenommen (P_n) ist straff, aber $P_n \not\Rightarrow P$.
Dann gibt es $\varepsilon > 0$, $f \in C_b(E)$ und eine Teilfolge, so dass

$$\left| E_{P_{n_k}}[f] - E_P[f] \right| > \varepsilon \quad \text{für alle } k. \tag{129}$$

$\Rightarrow$ Nach dem Satz von Prohorov gibt es eine Teilfolge $(P_{n_{k_l}})$ der straffen Familie $(P_{n_k})_{k\geq 1}$, so dass $P_{n_{k_l}} \overset{l\to\infty}{\Rightarrow} Q \in \mathscr{M}(E)$.
Dann ist

$$|E_P[f] - E_Q[f]| = \lim_{l\to\infty} \Big(E_P[f] - E_{P_{n_{k_l}}}[f] \Big) \geq \varepsilon$$

nach (129), also $P \neq Q$. Andererseits gilt nach Voraussetzung $\forall g \in \mathcal{M}$, dass

$$E_P[g] = \lim_{l\to\infty} E_{P_{n_{k_l}}}[g] = E_Q[g],$$

ein Widerspruch dazu, dass $\mathcal{M}$ separierend ist. □

Im unendlichdimensionalen Banachräumen sind abgeschlossene Kugeln nicht mehr kompakt (dies folgt aus dem Kompaktheitssatz von Riesz: ein normierter Vektorraum ist endlichdimensional genau dann, wenn $\overline{B}_1$ kompakt ist).

Der Satz von Arzela-Ascoli verknüpft Kompaktheit elegant mit folgendem Stetigkeitsbegriff:

Definition 130. Eine Familie von reellwertigen Funktionen $\mathcal{M}$ auf dem $\mathbb{R}^d$ heißt *gleichgradig stetig in* $x \in \mathbb{R}^d$, falls

$$\sup_{f\in\mathcal{M}} |f(x) - f(y)| \xrightarrow{y\to x} 0.$$

Dies kann man natürlich auch allgemein für Funktionen auf metrischen Räumen definieren. (**Gleichmäßige** Stetigkeit bezieht sich allerdings nur auf **eine** Funktion!)

Bemerkung 131. Handelt es sich um eine Folge (f_n) von stetigen Funktionen, so ist gleichgradige Stetigkeit äquivalent zu

$$\limsup_{n\to\infty} |f_n(x) - f_n(y)| \xrightarrow{y\to x} 0.$$

Wir erhalten folgendes Lemma:

Lemma 132. *Sei* $(f_n) \subseteq C(\mathbb{R}^d)$ *und gelte* $f_n(x) \to f(x)\ \forall x \in \mathbb{R}^d$. *Dann ist* (f_n) *genau dann gleichgradig stetig in* 0, *wenn* f *in* 0 *stetig ist.*

Beweis. Sei zunächst (f_n) gleichgradig stetig. Dann ist,

$$\begin{aligned} |f(x) - f(0) &= |\lim f_n(x) - \lim f_n(0)| = \lim |f_n(x) - f_n(0)| \\ &\leq \limsup |f_n(x) - f_n(0)| \to 0, \end{aligned}$$

da (f_n) gleichgradig stetig ist in 0.

Für die umgekehrte Richtung beobachten wir, dass

$$\begin{aligned}
&\limsup_{n\to\infty} |f_n(x) - f_n(0)| \\
&\qquad \le \limsup_{n\to\infty} \Big(|f_n(x) - f(x)| + |f(x) - f(0)| + |f(0) - f_n(0)| \Big) \\
&\qquad \overset{f_n(x)\to f(x)}{=} |f(x) - f(0)| \overset{x\to 0}{\to} 0,
\end{aligned}$$

da f stetig ist. $\square$

Wir benötigen noch eine handliche Abschätzung für die Fourier-Transformierte.

Lemma 133. *Für alle $r > 0$ und $P \in \mathscr{M}(\mathbb{R})$*

$$P\Big((-\infty, -r] \cup [r, \infty)\Big) \le \frac{r}{2} \int_{-2/r}^{2/r} (1 - \varphi_P(t))dt.$$

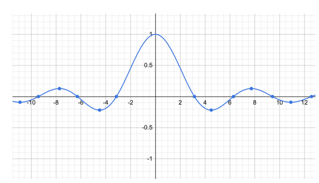
Beweis. Wir betrachten ein X mit Verteilung P. Dann ist

$$\int_{-c}^{c} (1 - \varphi_P(t)dt = E\Big[\int_{-c}^{c} (1 - e^{itX})dt\Big] = 2c - E\Big[\frac{1}{iX} e^{itX} \Big|_{-c}^{c}\Big]. \quad (134)$$

Da $e^{ix} = \cos x + i \sin x$ ist, und cos symmetrisch ist um Null, folgt $\cos t \,|_{-cX}^{cX} = 0$. Außerdem ist $\sin t \,|_{-cX}^{cX} = 2 \sin cX$, also ist

$$(134) \;= 2c\Big(E\Big[1 - \frac{\sin cX}{cX}\Big]\Big).$$

Nun ist $\frac{\sin x}{x} \begin{cases} > 0 & \text{für } |x| \le 2, \\ \le \frac{1}{2} & \text{für } |x| \ge 2. \end{cases}$.



Wir erhalten, dass $1 - \frac{\sin x}{x} \ge \frac{1}{2}$ für $|x| \ge 2 \Rightarrow$

$$\begin{aligned}
(134) \quad &\ge 2cE\Big[\Big(1 - \frac{\sin cX}{cX}\Big)\mathbb{1}_{\{|cX|\ge 2\}}\Big] \\
&\ge c \cdot P(|cX| \ge 2).
\end{aligned}$$

Die Behauptung folgt nun mit $c = \frac{2}{r}$. $\square$

Satz 135 (Arzelà-Ascoli)**.** *Sei $(P_i)_{i\in I} \subseteq \mathscr{M}(\mathbb{R}^d)$. (φ_{P_i}) ist gleichgradig stetig in 0 genau dann, wenn $(P_i)_{i\in I}$ straff ist.*

Beweis: Wir beweisen lediglich die eine Richtung: gleichgradige Stetigkeit an 0 impliziert Straffheit.

Es genügt zu zeigen, dass $(\pi_k P_i)_{i\in I}$ für alle Projektionen $\pi_1, \dots, \pi_d$ straff ist, und somit können wir ohne Einschränkung $d = 1$ annehmen[35].

Zunächst ist $\varphi_{P_i}(0) = 1$ für alle $i \in I$. Mit der Voraussetzung folgt, dass

$$\sup_{i\in I} |1 - \varphi_{P_i}(t)| \stackrel{t\to 0}{\to} 0.$$

Wir erhalten

$$\begin{aligned} \sup_{r>0} \inf_{i\in I} P_i([-r,r]) &\geq 1 - \inf_{r>0} \sup_{i\in I} \frac{r}{2} \int_{-2/r}^{2/r} (1 - \varphi_{P_i}(t))dt \\ &\geq 1 - \inf_{r>0} \frac{r}{2} \int_{-2/r}^{2/r} \sup_{i\in I} (1 - \varphi_{P_i}(t))dt \\ &\geq 1 - 2 \inf_{r>0} \sup_{[-2/r,2/r]} \sup_{i\in I} (1 - \varphi_{P_i}(t)) = 1. \qquad \square \end{aligned}$$

[35] Man beobachte, dass $\{x \in \mathbb{R}^d : x \notin [-r,r]^d\} = \cup_{1\leq k\leq d}\{x \in \mathbb{R}^d : x_k \notin [-r,r]\}$. Umgekehrt ist für $1 \leq k \leq d$ $\{x \in \mathbb{R}^d : x_k \in [-r,r]\} \supseteq \{x \in \mathbb{R}^d : x \in [-r,r]^d\}$.

Der Satz von Lévy besagt nun, dass falls punktweise Konvergenz der charakteristischen Funktionen vorliegt und diese stetig an 0 sind, so folgt bereits die Verteilungskonvergenz.

Theorem 136 (Lévy)**.** *Seien $(P_n) \subseteq \mathscr{M}(\mathbb{R}^d)$ und $\varphi_{P_n}(t) \to \varphi(t)$ für alle $t \in \mathbb{R}^d$. Ist φ stetig in 0, so folgt $P_n \Rightarrow P$ für ein $P \in \mathscr{M}(\mathbb{R}^d)$ und $\varphi = \varphi_P$.*

Beweis. Da φ_{P_n} punktweise konvergieren und φ in 0 stetig ist, folgt mit Lemma 132, dass (φ_{P_n}) in 0 gleichgradig stetig ist. Mit dem Satz von Arzelà-Ascoli, Satz 135, folgt also, dass (P_n) straff ist, also relativ folgenkompakt. Nach dem Satz von Prohorov, Satz 115, wählen wir eine Teilfolge $P_{n_k} \Rightarrow P \in \mathscr{P}(\mathbb{R}^d)$.

Da die Abbildung $x \mapsto e^{itx}$ stetig und beschränkt ist, folgt

$$\varphi_{P_{n_k}}(t) \to \varphi_P(t) \quad \text{für alle } t \in \mathbb{R}^d.$$

Nach Voraussetzung gilt auch $\varphi_{P_n}(t) \to \varphi_P(t)$, also $\varphi = \varphi_P$. $\square$

Beispiel 137 (Satz von de Moivre-Laplace)**.** Ist S_n binomialverteilt, $S_n \sim B(n,p)$, so konvergiert die standardisierte Zufallsvariable S_n^* gegen eine Standard-Normalverteilung, also $S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}} \Rightarrow \mathcal{N}(0,1)$:

Zunächst ist $S_n = \sum_{i=1}^n X_i$ mit $X_i \sim \text{Bin}(1,p)$, also

$$\begin{aligned} \varphi_{S_n}(t) &= \big(pe^{it} + (1-p)\big)^n \\ \text{und} \quad \varphi_{S_n/\sqrt{np(1-p)}}(t) &= \big((1-p) + pe^{it/\sqrt{np(1-p)}}\big)^n, \end{aligned}$$

also

$$\varphi_{S_n^*} > e^{-it\sqrt{\frac{np}{1-p}}}\Big((1-p)+pe^{\frac{it}{\sqrt{np(1+p)}}}\Big)^n$$
$$= \Big((1-p)e^{it\sqrt{\frac{p}{n(1-p)}}} + pe^{\frac{it}{\sqrt{np(1-p)}}-it\sqrt{\frac{p}{n(1-p)}}}\Big)$$
$$\overset{\text{Taylor}}{=} \Big((1-p)\Big(1-it\sqrt{\frac{p}{n(1-p)}}-\frac{1}{2}\frac{t^2}{n(1-p)}-\underbrace{\frac{1}{b}it^3\Big(\frac{p}{n(1-p)}\Big)^{3/2}}_{\to O\left(\frac{1}{n^{3/2}}\right)}\Big)$$
$$+p\Big(1+\frac{it\sqrt{1-p}}{\sqrt{np}}-\frac{1}{2}\frac{t^2(1-p)}{np}+\underbrace{\frac{1}{b}it^3\Big(\frac{p}{n(1-p)}\Big)^{3/2}}_{\to O\left(\frac{1}{n^{3/2}}\right)}\Big)\Big)^n$$
$$= \Big(1-\underbrace{\frac{t^2}{2}\big(\frac{p}{n}+\frac{1-p}{n}\big)}_{=-\frac{t^2}{2n}}+C_n\cdot\frac{1}{n^{3/2}}\Big)^n \to e^{-\frac{t^2}{2}} = \varphi_{N(0,1)}(t).$$

Für die Laplace-Transformierte erhält man eine analoge Aussage.

Beispiel 138. Gilt $X_n \sim$Geom(p_n) und $n\cdot p_n \to \lambda$, so folgt $\frac{X_n}{n} \Rightarrow$Exp$(\lambda)$:
Wir erhalten, dass

$$E\Big[e^{it\frac{X_n}{n}}\Big] = \sum_{k=1}^{\infty}(1-p_n)^{k-1}\cdot p_n e^{\frac{itk}{n}}$$
$$= \frac{p_n}{1-p_n}\cdot\Big(\sum_{k=0}^{\infty}\Big((1-p_n)\cdot e^{\frac{it}{n}}\Big)-1\Big) \tag{139}$$

Nun ist

$$\sum_{k=0}^{\infty}\Big((1-p_n)\cdot e^{\frac{it}{n}}\Big)-1$$
$$= \frac{1}{1-(1-p_n)e^{\frac{it}{n}}} - \frac{1-(1-p_n)\cdot e^{\frac{it}{n}}}{1-(1-p_n)\cdot e^{\frac{it}{n}}} = (1-p_n)\cdot\frac{e^{\frac{it}{n}}}{1-(1-p_n)\cdot e^{\frac{it}{n}}}.$$

Damit erhalten wir

$$(139) = \frac{np_ne^{\frac{it}{n}}}{n(1-(1-p_n)e^{\frac{it}{n}})} \to \frac{\lambda}{\lambda-it} = \varphi_{\mathrm{Exp}(\lambda)},$$

da

$$n(e^{-\frac{it}{n}}-(1-p_n)) = n\Big(1-\frac{it}{n}+O\Big(\frac{1}{n^2}\Big)-1+p_n\Big) = -it+n\cdot p_n+O\Big(\frac{1}{n}\Big).$$

Der zentrale Grenzwertsatz von Lindeberg und Feller

Der zentrale Grenzwertsatz von Lindeberg (1922) und Feller (1935, Theorem 140, verallgemeinert den Satz von de Moivre und Laplace sowie den zentralen Grenzwertsatz von Lindeberg und Lévy. Hierbei wird die Annahmen von unabhängig und identischer Verteilung durch die Annahme von Unabhängigkeit in Kombination mit der *Lindeberg-Bedingung* (LB) ersetzt. Lindeberg zeigte hierbei zunächst, aus (ii) folgt (i), während Feller die Rückrichtung bewies.

Theorem 140 (Zentraler Grenzwertsatz von Lindeberg-Feller)**.**
Sei $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,m_n}$ *eine Familie von Zufallsvariablen, so dass für* $n = 1, 2, \ldots$ *die Zufallsvariablen* $X_{n1}, \ldots, X_{nm_n}$ *unabhängig sind. Sei außerdem*

$$\sum_{j=1}^{m_n} E[X_{nj}] \xrightarrow{n\to\infty} \mu, \qquad \sum_{j=1}^{m_n} \mathrm{Var}(X_{nj}) \xrightarrow{n\to\infty} \sigma^2$$

und $X \sim \mathcal{N}(\mu, \sigma^2)$*. Dann sind folgende Aussagen äquivalent:*

(i) $\displaystyle\sum_{j=1}^{m_n} X_{nj} \xRightarrow{n\to\infty} X$ *und* $\displaystyle\sup_{j=1,\ldots,m_n} \mathrm{Var}[X_{nj}] \xrightarrow{n\to\infty} 0,$

(ii) $\displaystyle\sum_{j=1}^{m_n} E\Big[(X_{nj} - E[X_{nj}])^2 \mathbb{1}_{\{|X_{nj}-E[X_{nj}]|>\varepsilon\}}\Big] \xrightarrow{n\to\infty} 0$ *für alle* $\varepsilon > 0.$
(LB)

Als einfache Folgerund erhalten wir den zentralen Grenzwertsatz für i.i.d. Zufallsvariablen, den wir bereits in der Stochastik I kennengelernt hatten.

Korollar 141. *Seien* $X_1, X_2, \ldots$ *i.i.d. mit* $E[X_1] = \mu, \mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$*. Dann gilt*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \xRightarrow{n\to\infty} \mathcal{N}(0,1).$$

Beweis. Sei $m_n = n$ und $X_{nj} = \frac{X_n - \mu}{\sqrt{n\sigma^2}}$. Dann erfüllt die Familie $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,n}$ die Voraussetzungen von Theorem 140 mit $\mu = 0, \sigma^2 = 1$. Außerdem gilt

$$\sum_{j=1}^{n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] = \frac{1}{\sigma^2} E[(X_1 - \mu)^2 \mathbb{1}_{\{|X_1-\mu|>\varepsilon\sqrt{n\sigma^2}\}}] \xrightarrow{n\to\infty} 0$$

mit majorisierter Konvergenz. □

Oftmals ist die Lindeberg-Bedingung nicht einfach nachzuprüfen. Einfacher ist oft die stärkere[36] *Lyapunov-Bedingung*.

[36] Aktuelle deutsche Schreibweise nach Wikipedia ist Lyapunow, im Englischen of Lyapunov.

Bemerkung 142. Die *Lyapunov-Bedingung* gilt im Kontext von Theorem 140, falls für ein $\delta > 0$

$$\sum_{j=1}^{m_n} E\big[|X_{nj} - E[X_{nj}]|^{2+\delta}\big] \xrightarrow{n\to\infty} 0. \tag{143}$$

Die Lyapunov-Bedingung impliziert die Lindeberg-Bedingung: Ohne Einschränkung sei $E[X_{nj}] = 0$. Es gilt für $\varepsilon > 0$

$$x^2 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta} 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta}.$$

Gilt nun die Lyapunoff-Bedingung, so folgt die Lindeberg-Bedingung aus

$$\sum_{j=1}^{m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \leq \frac{1}{\varepsilon^\delta} \sum_{j=1}^{m_n} E[|X_{nj}|^{2+\delta}] \xrightarrow{n\to\infty} 0. \qquad \diamond$$

Der Beweis von Theorem 140 basiert auf der geschickten Verwendung der charakteristischen Funktionen der Zufallsvariable X_{nj} und Taylor-Approximationen. Wir bereiten den Beweis des Theorems mit zwei Lemmata vor.

Lemma 144. *Für komplexe Zahlen $z_1, \ldots, z_n, z'_1, \ldots, z'_n$ mit $|z_i| \leq 1, |z'_i| \leq 1$ für $i = 1, \ldots, n$ gilt*

$$\Big| \prod_{k=1}^n z_k - \prod_{k=1}^n z'_k \Big| \leq \sum_{k=1}^n |z_k - z'_k|. \tag{145}$$

Beweis. Die Idee für den Beweis ist recht simpel: Wir haben

$$z_1 z_2 - z'_1 z'_2 = z_1(z_2 - z'_2) + (z_1 - z'_1) z'_2.$$

Fügt man nun Beträge hinzu, kann man $|z_1|$ und $|z'_2|$ durch 1 abschätzen und die Ungleichung folgt für $n = 2$. Der Rest ist Induktion: Für $n = 1$ ist die Gleichung offensichtlich richtig. Gilt (145) für ein n, so ist

$$\begin{aligned}\Big| \prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} z'_k \Big| &\leq \Big| z_{n+1} \Big(\prod_{k=1}^n z_k - \prod_{k=1}^n z'_k \Big) \Big| + \Big| (z_{n+1} - z'_{n+1}) \prod_{k=1}^n z'_k \Big| \\ &\leq \sum_{k=1}^n |z_k - z'_k| + |z_{n+1} - z'_{n+1}|.\end{aligned}$$

Daraus folgt die Behauptung. $\square$

Wir verwenden weiterhin folgende Taylor-Approximation der Exponentialfunktion.

Lemma 146. *Sei $t \in \mathbb{C}$ und $n \in \mathbb{Z}_+$. Dann gilt*

$$\Big| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \Big| \leq \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}. \tag{147}$$

Beweis. Die zentrale Idee für die Approximation ist

$$e^{it} - 1 = i \int_0^t e^{is} ds$$

zu nutzen in Kombination mit Jensen. So folgt

$$|e^{it} - 1| = \Big| \int_0^t e^{is} ds \Big| \leq \int_0^t |e^{is}| ds = |t|.$$

Außderdem haben wir natürlich

$$|e^{it} - 1| \leq 1 + 1 = 2.$$

Wir bezeichnen mit $h_n(t)$ die Differenz auf der linken Seite von (147). Für $n = 0$ haben wir die Behauptung bereits gezeigt. Allgemein gilt für $t \in \mathbb{R}, n \in \mathbb{N}$

$$\Big| \int_0^t h_n(s) ds \Big| = \Big| -i(e^{it} - 1) + i \sum_{k=0}^{n} \frac{(it)^{k+1}}{(k+1)!} \Big| = \Big| ie^{it} - i \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \Big| = |h_{n+1}(t)|.$$

Mittels Induktion erhalten wir nun

$$|h_{n+1}(t)| \leq \Big| \int_0^t \frac{2|s|^n}{n!} ds \Big| = \frac{2|t|^{n+1}}{n!(n+2)}, \text{ und}$$
$$|h_{n+1}(t)| \leq \Big| \int_0^t \frac{|s|^{n+1}}{(n+1)!} ds \Big| = \frac{|t|^{n+2}}{(n+1)!(n+2)}.$$

woraus (147) folgt. □

Lemma 148. *Für eine Zufallsvariable X mit $E[X] = 0$ und $\operatorname{Var}(X) = \sigma^2 > 0$ ist für jedes $\varepsilon > 0$*

$$E[X^2 \mathbb{1}_{\{|x| \geq \varepsilon\}}] \leq \varepsilon^2 \Big(Re\varphi_X(\sqrt{6}/\varepsilon) - 1 + 3\frac{\sigma^2}{\varepsilon^2} \Big). \tag{149}$$

Beweis. Hierbei ist $\mathrm{Re}\varphi_X(t) = E[\cos(tX)]$ und wir nutzen die einfachen Abschätzungen $\cos x \leq 1$ sowie $1 - \cos(x) \leq \frac{x^2}{2}$.

Damit ist

$$\begin{aligned}
\mathrm{Re}\varphi_X(t) - 1 + \frac{\sigma^2 t^2}{2} &= \frac{\sigma^2 t^2}{2} - \int (1 - cos(tx)) F_X(dx) \\
&= \frac{\sigma^2 t^2}{2} - \int_{|x|<\varepsilon} (1 - cos(tx)) F_X(dx) - \int_{|x|\geq\varepsilon} (1 - cos(tx)) F_X(dx) \\
&\geq \frac{\sigma^2 t^2}{2} - \frac{t^2}{2} \int_{|x|<\varepsilon} x^2 F_X(dx) - \frac{2}{\varepsilon^2} \int_{|x|\geq\varepsilon} x^2 F_X(dx) \\
&= \Big(\frac{t^2}{2} - \frac{2}{\varepsilon^2} \Big) \int_{|x|\geq\varepsilon} x^2 F_X(dx)
\end{aligned}$$

und die Behauptung folgt mit $t = \sqrt{6}\varepsilon^{-1}$. □

Im nun folgenden Beweis werden wir für Funktionen f und g $f \lesssim g$ schreiben, falls es eine Konstante C gibt mit $f \leq Cg$.

Beweis (Theorem 140). Ohne Einschränkung sei $E[X_{nj}] = \mu = 0$ und $\sigma^2 = 1$. Wir setzen

$$\begin{aligned} \sigma_{nj}^2 &= \mathrm{Var}[X_{nj}], \\ \sigma_n^2 &= \sum_{j=1}^{m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} \sigma^2 = 1, \\ \varphi_{nj}(t) &= E[e^{itX_{nj}}]. \end{aligned}$$

Wir zeigen zunächst den Satz von Lindeberg: aus der Lindeberg-Bedingung (ii) folgt (i): Da für jedes $\varepsilon > 0$

$$\sup_{j=1,\dots,m_n} \sigma_{nj}^2 \le \varepsilon^2 + \sup_{j=1,\dots,m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \le \varepsilon^2 + \sum_{j=1}^{m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \xrightarrow{n\to\infty} \varepsilon^2, \tag{150}$$

folgt $\sup_{j=1,\dots,m_n} \mathrm{Var}[X_{nj}] \xrightarrow{n\to\infty} 0$, also ist der zweite Teil von (i) bereits gezeigt.

Seien $(Z_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ unabhängige Zufallsvariablen mit $Z_{nj} \sim N(0, \sigma_{nj}^2)$. Dann ist $Z_n := \sum_{j=1}^{m_n} Z_{nj} \sim N(0, \sigma_n^2)$. Insbesondere gilt also $Z_n \xRightarrow{n\to\infty} X$, was man etwa direkt aus der Form der charakteristischen Funktion der Normalverteilung abliest.

Sei $\widetilde{\varphi}_{nj} = e^{-\frac{1}{2}\sigma_{nj}^2 t^2}$ die charakteristische Funktion von Z_{nj}. Dann genügt es zu zeigen, siehe Theorem 136, dass

$$\prod_{j=1}^{n_j} \varphi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\varphi}_{nj}(t) \xrightarrow{n\to\infty} 0 \tag{151}$$

für alle t. Mittels Lemma 144 und Lemma 146 schreiben wir

$$\begin{aligned} \Big| \prod_{j=1}^{m_n} \varphi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\varphi}_{nj}(t) \Big| &\le \sum_{j=1}^{m_n} |\varphi_{nj}(t) - \widetilde{\varphi}_{nj}(t)| \\ &= \sum_{j=1}^{m_n} |\varphi_{nj}(t) - 1 + \frac{1}{2}t^2\sigma_{nj}^2 - \big(\widetilde{\varphi}_{nj}(t) - 1 + \frac{1}{2}t^2\sigma_{nj}^2\big)| \\ &\le \sum_{j=1}^{m_n} |\varphi_{nj}(t) - 1 + \tfrac{1}{2}t^2\sigma_{nj}^2| + \sum_{j=1}^{m_n} |\widetilde{\varphi}_{nj}(t) - 1 + \tfrac{1}{2}t^2\sigma_{nj}^2|. \end{aligned}$$

Nun haben wir durch Lemma 146, dass

$$|\varphi_{nj}(t) - 1 + \frac{1}{2}t^2\sigma_{nj}^2| \le E\left[\frac{2|it|\,|X_{nj}|^2}{2!} \wedge \frac{|it|\,|X_{nj}|^3}{3!}\right] \lesssim E[X_{nj}^2(1 \wedge |X_{nj}|)].$$

Weiter ist durch Einfügen von $\mathbb{1}_{\{|X_{nj}|\le\varepsilon\}} + \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}$

$$\sum_{j=1}^{m_n} E[X_{nj}^2(1 \wedge |X_{nj}|)] \le \varepsilon \sum_{j=1}^{m_n} \sigma_{nj}^2 + \sum_{j=1}^{m_n} E[X_{nj}^2 \mathbb{1}_{\{|X_{nj}|>\varepsilon\}}] \xrightarrow{n\to\infty} \varepsilon.$$

Nutzen wir Lemma 146 mit $n = 1$, so erhalten wir

$$|e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \tfrac{1}{2}t^2\sigma_{nj}^2| \leq \frac{(\frac{1}{2}\sigma_{nj}^2 t^2)^2}{2!} \leq \frac{t^4}{8}\sigma_{nj}^4,$$

also

$$\sum_{j=1}^{m_n} |e^{-\frac{1}{2}\sigma_{nj}^2 t^2} - 1 + \tfrac{1}{2}t^2\sigma_{nj}^2| \lesssim \sum_{j=1}^{m_n} \sigma_{nj}^4 \leq \sigma_n^2 \sup_{j=1,\dots,m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 0$$

wegen (150). Damit ist (151) bereits bewiesen.
Nun zeigen wir den Satz von Feller, (i) impliziert (ii). Nach dem zweiten Teil von (i) ist für jedes $\varepsilon > 0$ mit der Tschebyscheff-Ungleichung

$$\sup_{j=1,\dots,m_n} P[|X_{nj}| > \varepsilon] \leq \sup_{j=1,\dots,m_n} \frac{\sigma_{nj}^2}{\varepsilon^2} \xrightarrow{n\to\infty} 0. \tag{152}$$

Mit Lemma 146 gilt

$$\sup_{j=1,\dots,m_n} |\varphi_{nj}(t) - 1| \leq \sup_{j=1,\dots,m_n} E[2 \wedge |t \cdot X_{nj}|] \leq 2 \sup_{j=1,\dots,m_n} P[|X_{nj}| > \varepsilon] + \varepsilon|t| \xrightarrow{n\to\infty} \varepsilon|t|.$$

Insbesondere ist $\sum_{j=1}^{m_n} \log \varphi_{nj}(t)$ für jedes t definiert, falls n groß genug ist. Aus der Gültigkeit von (i) folgt

$$\sum_{j=1}^{m_n} \log \varphi_{nj}(t) \xrightarrow{n\to\infty} -\frac{t^2}{2}. \tag{153}$$

Außerdem gilt wegen $\varphi'_{nj}(0) = iE[X_{nj}] = 0$, $\varphi''_{nj}(0) = -\operatorname{Var}[X_{nj}] = -\sigma_{nj}^2$ mit Hilfe einer Taylorentwicklung von φ_{nj} um 0 (Lemma 146)

$$|\varphi_{nj}(t) - 1| \lesssim \sigma_{nj}^2 |t|^2$$

und

$$\Big| \sum_{j=1}^{m_n} \log \varphi_{nj}(t) - \sum_{j=1}^{m_n} (\varphi_{nj}(t) - 1) \Big| \lesssim \sum_{j=1}^{m_n} |\varphi_{nj}(t) - 1|^2 + \dots$$

mit Termen höherer Ordnung. Nun gilt aber,

$$\sum_{j=1}^{m_n} |\varphi_{nj}(t) - 1|^2 \lesssim \sum_{j=1}^{m_n} (\sigma_{nj}^2)^2 |t|^4 \leq |t|^4 \sup_{j=1,\dots,m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 0 \tag{154}$$

und somit verschwinden auch die Terme höherer Ordnung.

Da aus der Konvergenz einer imaginären Reihe die Konvergenz ihres Realteils folgt, folgern wir aus (153) und (154) wegen $\operatorname{Re}(\varphi_{nj}(t)) = E[\cos(tX_{nj})]$

$$\sum_{j=1}^{m_n} E[\cos(tX_{nj}) - 1] \xrightarrow{n\to\infty} -\frac{t^2}{2}.$$

Da $\sum_{j=1}^{m_n} \sigma_{nk}^2 \to 1$ erhalten wir

$$\sum_{j=1}^{m_n} \left(E[\cos(tX_{nj})] - 1 + \frac{t^2\sigma_{nk}^2}{2} \right) \to 0 \tag{155}$$

für alle t. Aus dem Lemma 148 folgt damit, dass

$$\sum_{j=1}^{m_n} E[X_{nj}^2(1 - \mathbb{1}_{\{|X_{nj}| \le \varepsilon\}})] \le \varepsilon^2 \sum_{j=1}^{m_n} \left(E[\cos(\sqrt{6}/\varepsilon X_{nk})] - 1 + 3\frac{\sigma_{nk}^2}{\varepsilon^2} \right) \to 0$$

aus (155) mit $t = \sqrt{6}/\varepsilon$ und somit (ii). □

Konvergenz des Binomialmodells gegen das Black-Scholes Modell

Zwei wichtige Modelle für Aktienkurse sind das Binomialmodell und das Black-Scholes Modell. Für Literatur verweisen wir auf [37].

[37] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter, Berlin, 2011

Zentrales Hilfsmittel wird ein zentraler Grenzwertsatz unter Beschränktheitsannahmen sein, welchen wir als erstes betrachten.

Theorem 156. *Seien $Y_{1n}, \dots, Y_{nn}$ unabhängig für alle $n \ge 1$ mit*

(i) $|Y_{kn}| \le \gamma_n$ *P-f.s. mit* $\gamma_n \longrightarrow 0$,

(ii) $\sum_{k=1}^{n} E[Y_{kn}] \longrightarrow \mu$,

(iii) $\sum_{k=1}^{n} \mathrm{Var}(Y_{kn}) \longrightarrow \sigma^2 < \infty$.

Dann gilt, dass

$$\sum_{k=1}^{n} Y_{kn} \xrightarrow{\mathscr{L}} \mathscr{N}(\mu, \sigma^2).$$

Für den Beweis des Theorems weißt man die Lyapunov-Bedingung (143) nach.

Als nächstes führen wir das Binomialmodell ein, zunächst für ein festes n. Der Markt besteht aus einem Bankkonto mit Preisprozess S^0 und einer Aktie mit Preisprozess S^1. Als Zeitskala betrachten wir $0, 1, \dots, n$. Das Bankkonto entwickelt sich mit dem Zinssatz, so dass

$$S_t^0 = S_{tn}^0 = (1 + r_n)^t.$$

Darüber hinaus gilt für die Aktie ein multiplikatives Modell,

$$S_t = S_{tn} = S_{t-1}(1 + R_{tn}).$$

Hierbei ist $R_t = R_{tn}$ der Return oder die Rendite der Periode t. Wir betrachten $\Omega = \{-1, 1\}^n$ und setzen

$$R_{tn} = \begin{cases} a_n & \text{falls } \omega_t = -1 \\ b_n & \text{falls } \omega_t = 1. \end{cases}$$

Handelsstrategien

Die Bank kann nun selbstfinanzierend in der Akte handeln. Selbstfinanzierend heißt, dass in der Zeitperiode $(0, n)$ weder Geld abgezogen noch zugeschossen wird. Besonders leicht lässt sich dies in diskontierten Größen betrachten und wir setzen

$$X_t = \frac{S_t}{S_t^0},$$

der diskontierte Preisprozess. Eine selbstfinanzierende Handelsstrategie wird durch einen *adaptierten* Prozess H dargestellt - die (diskontierten) Gewinne am Ende der Periode ergeben sich zu

Diese Begrifflichkeit werden wir in kürze klären

$$(H \cdot X)_n := \sum_{t=0}^{n-1} H_t \Delta X_t = \sum_{t=0}^{n-1} H_t (X_{t+1} - X_t).$$

Der zugehörige (diskontierte) Wertprozess $V = V^H$ ist gegeben durch $V_0 = H_0 \cdot S_0$ und $V_n = V_0 + (H \cdot H)_n$

Eine Arbitrage ist eine selbstfinanzierende Handelsstrategie H, so dass

$$V_0 \leq 0, \qquad V_n \geq 0, P(V_n > 0) > 0.$$

Zentrales Hilfsmittel in der Finanzmathematik ist folgender Hauptsatz der Wertpapierbewertung:

Theorem 157. *Der Finanzmarkt ist frei von Arbitrage, falls ein äquivalentes Martingalmaß P^* existiert.*

Einen Beweis werden wir in Kürze führen können (zumindest für die in der Praxis relevante Rückrichtung !)

Wir nehmen nun noch an, dass das statistische Maß P gegeben ist durch

$$P(\omega) = \prod_{t=1}^{n} p(\omega_t).$$

Für das äquivalente Martingalmaß P^* gilt sogar, dass $(R_{tn})_{t=1,\dots,n}$ unabhängig sind und die Martingalebedingung erfüllt ist. Diese erhält man unter

$$p^* = \frac{r_n - a_n}{b_n - a_n},$$

denn dann gilt, dass

$$E^*\Big[\frac{1}{(1+r_n)^n} S_{tn}\Big] = S_0.$$

Für die Konvergenz machen wir die folgenden Annahmen mit einem festen Zeithorizont $T > 0$:

(i) $\lim_{n\to\infty}(1+r_n)^n = e^{rT}$,
(ii) $-1 < \alpha_n \leq R_{tn} \leq \beta_n$ mit $\alpha_n, \beta_n \to 0$,
(iii) $\sum_{t=1}^n \operatorname{Var}^*(R_{tn}) \to \sigma^2 T \in (0,\infty)$.

Theorem 158. *Unter (i) - (iii) gilt, dass*

$$S_{nn} \xrightarrow{\mathscr{L}} S_T = S_0 \exp\Big(\sigma W_T + (r - \frac{\sigma^2}{2})T\Big),$$

mit $W_T \sim \mathcal{N}(0,T)$.

Die Log-Normalverteilung auf der rechten Seite der obigen Gleichung spezifiziert exakt das berühmte *Black-Scholes Modell*[38].

[38] Siehe die mit einem Nobelpreis gewürdigte Arbeit

Fisher Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973

Beweis. Ohne Einschränkung sei $S_0 = 1$. Wir nutzen folgende Taylor-Entwicklung:

$$\log(1+x) = x - \frac{x^2}{2} + \rho(x)\cdot x^2,$$

mit $|\rho(x)| \leq \delta(\alpha,\beta)$ und $-1 < \alpha \leq x \leq \beta$. Hierbei gilt $\delta(\alpha,\beta) \to 0$ falls $\alpha,\beta \to 0$. Wir erhalten, dass

$$\log(S_n^n) = \sum_{k=1}^n \log(1+R_k^n) = \sum_{k=1}^n \Big(R_k^n - \frac{(R_k^n)^2}{2}\Big) + \Delta_n$$

mit $|\Delta_n| \leq \delta(\alpha_n,\beta_n)\cdot\sum_{k=1}^n (R_k^n)^2$. Unter der Wahl von p^* ist

$$E^*[R_k^n] = r_k^n,$$

so dass

$$E^*[|\Delta_n|] \leq \delta(\alpha_n,\beta_n)\cdot\sum_{k=1}^n \Big(\operatorname{Var}(R_k^n) + (r_n^n)^2)\Big) \to 0;$$

der erste Teil der Summe konvergiert nach Annahme gegen σ^2, der zweite Teil gegen r^T. Es folgt, dass $\Delta_n \xrightarrow{P} 0$, und $\Delta_n \xrightarrow{\mathscr{L}} \delta_0$. Es bleibt zu zeigen, dass (Slutsky)

$$Z_n = \sum_{k=1}^n \Big(\underbrace{R_k^n - \frac{1}{2}(R_k^n)^2}_{=:Y_k^n}\Big) \Rightarrow \mathcal{N}(rT - \frac{\sigma^2 T}{2}, \sigma^2 T).$$

Zunächst ist $\max|Y_k^n| \leq \gamma_n + \frac{1}{2}\gamma_n^2 \to 0$ mit $\gamma_n = \max\{|\alpha_n|,|\beta_n|\}$. Wir erhalten

$$E[Z_n] = n\cdot r_n - \frac{1}{2}(\sigma_n^2 T + nr_n^2) \to rT - \frac{\sigma^2 T}{2},$$
$$\operatorname{Var}(Z_n) \to \sigma^2 T,$$

da

$$\operatorname{Var}(Z_n) \leq \sum_{k=1}^{n} \operatorname{Var}(R_k^n) + \frac{n}{4} \operatorname{Var}((R_k^n)^2),$$

$$\sum_{k=1}^{n} E[|R_k^n|^p] \leq \gamma_n^{p-2} \sum_{k=1}^{n} E[(R_k^n)^2].$$

□

Mehrdimensionale Grenzwertsätze

Bisher haben wir schwache Grenzwertsätze nur für den Fall $\mathbb{R}$–wertiger Zufallsvariablen betrachtet. Wir verallgemeinern dies nun zu $\mathbb{R}^d$–wertigen Zufallsgrößen. Insbesondere geben wir eine Variante des mehrdimensionalen zentralen Grenzwertsatzes an. Einen Vektor im $\mathbb{R}^d$ fassen wir als Spaltenvektor auf[39]. Das Skalarprodukt schreibt sich dann als $\langle a, b\rangle = a^\top b$.

[39] Im Skriptum von Prof. Pfaffelhuber sind dies gerade Zeilenvektoren.

Strikt positiv definit bedeutet $x^\top C x > 0$ für alle $x \in \mathbb{R}^d$. Aus der linearen Algebra ist bekannt, dass es für eine strikt positiv definite Matrix C immer eine invertierbare Matrix A gibt mit $C = AA^\top$.

Definition 159. Seien $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$ eine strikt positiv definite symmetrische Matrix. Die *d-dimensionale Normalverteilung* mit Erwartungswert μ und Covarianzmatrix Σ hat die Dichte

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\Big(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big).$$

Satz 160. *Seien $\mu \in \mathbb{R}^d$, $\Sigma = AA^\top \in \mathbb{R}^{d\times d}$ strikt positiv definit und symmetrisch. Dann sind äquivalent:*

(i) $X \sim \mathcal{N}(\mu, \Sigma)$
(ii) $\langle t, X\rangle \sim \mathcal{N}(\langle t, \mu \rangle, t^\top \Sigma t)$ *für jedes* $t \in \mathbb{R}^d$
(iii) $\varphi_X(t) = \exp\Big(i\langle t, \mu\rangle - \tfrac{1}{2} t^\top \Sigma t\Big)$ *für jedes* $t \in \mathbb{R}^d$.

In jedem dieser Fälle gilt

(iv) $X \stackrel{\mathcal{L}}{=} AY + \mu$ *für* $Y \sim \mathcal{N}(0, Id_d)$
(v) $E[X_i] = \mu_i$ *für* $i = 1, \dots, d$
(vi) $\operatorname{Cov}(X_i, X_j) = \Sigma_{ij}$ *für* $i, j = 1, \dots, d$

Damit erhalten wir folgende Version des mehrdimensionalen zentralen Grenzwertsatzes:

Theorem 161. *Seien $X_1, X_2, \ldots$ unabhängige, identische verteilte Zufallsvariablen mit Werten in $\mathbb{R}^d$ mit $E[X_n] = \mu \in \mathbb{R}^d$ und $\mathrm{Cov}[X_{n,i}, X_{n,j}] = \Sigma_{ij}$ für $i, j = 1, \ldots, d$. Dann gilt*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{n \to \infty} \mathcal{N}_{0,\Sigma}.$$

Beweis. Wir wenden den eindimensionalen zentralen Grenzwertsatz, Korollar 141, auf die unabhängigen, identisch verteilten Zufallsvariablen $tX_1, tX_2, \ldots$ an. Dieser liefert

$$t^\top \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{n \to \infty} t^\top X.$$

Da t beliebig war, folgt die Aussage mit Hilfe des Cramer-Wold Device. $\square$

Martingale

In diesem Abschnitt wird das äußerst nützliche Hilfsmittel Martingale eingeführt und die Konvergenzaussagen aus dem letzten Abschnitt damit deutlich über die i.i.d.-Situation hinaus verallgemeinert.

Bedingte Erwartung

In Stochastik I haben wir bereits bedingte Erwartungswerte im Fall beschrieben, wenn $\mathscr{F} = (F_i)_{i \in I}$ eine abzählbare Partition von Ω ist:

$$E[X \mid \mathscr{F}] = \sum_{i \in I : P(F_i) > 0} \mathbb{1}_{F_i}(\omega) E[X \mid F_i].$$

Damit ist $E[X|\mathscr{F}]$ eine Zufallsvariable. Gegenstand dieses Abschnittes ist es, diese Definition für beliebige σ-Algebren $\mathscr{F}$ einzuführen.

Wir betrachten einen Wahscheinlichkeitsraum $(\Omega, \mathscr{A}, P)$.

Definition 162. Ist $X \in L^1(P)$ eine Zufallsvariable und $\mathscr{F} \subseteq \mathscr{A}$ eine Sub-σ-Algebra, so heißt Y *bedingte Erwartung von X gegeben $\mathscr{F}$*, falls

(i) Y ist $\mathscr{F}$-messbar

(ii) $\int_F Y dP = \int_F X dP$ für alle $F \in \mathscr{F}$

Wir schreiben $E[X \mid \mathscr{F}]$ für ein jedes solches Y und betrachten den bedingten Erwartungswert als Äquivalenzklasse dieser Zufallsvariablen. Diese sind P-f.s. gleich.

Satz 163. *Zu jedem $X \in L^1(P)$ und jeder Sub-σ-Algebra $\mathscr{F}$ existiert ein*

$$Y = E[X \mid \mathscr{F}] \in L^1(P).$$

Ist Y' ebenfalls eine bedingte Erwartung von X gegeben $\mathscr{F}$, so ist

$$P(Y = Y') = 1.$$

Beweis. Für die Existenz nutzen wir den Satz von Radon-Nikodým, Satz 61. Wir definieren

$$\mu(F) = \int_F X dP.$$

Dann ist μ ein signiertes Maß auf $(\Omega, \mathscr{F})$ und offensichtlich $\mu \ll P|_{\mathscr{F}}$. Demnach existiert eine $\mathscr{F}$-messbare Dichte Y, so dass

$$\nu(F) = \int_F Y dP, \tag{164}$$

und somit erfüllt Y Punkt (ii) von Definition 162.

Für (i) dieser Definition betrachten wir nun $Y = E[X \,|\, \mathscr{F}]$ und $F = \{Y > 0\}$, so gilt, da $F \in \mathscr{F}$,

$$\int_F Y dP = \int_F X dP \leq \int_F |X| dP,$$
$$\int_{F^c} -Y dP = \int_{F^c} -X dP \leq \int_{F^c} |X| dP,$$

also $E[|Y|] \leq E[|X|] < \infty$, und somit $Y \in L^1(P)$.

Sind weiterhin Y und Y' bedingte Erwartungen und $F = \{Y' - Y > 0\} \in \mathscr{F}$. Dann ist

$$\int_F (Y - Y') dP = \int_F (X - X') dP = 0$$

und ebenso für F^c, also $Y = Y' \quad P$-f. s. □

Satz 165. *Für $X, Y \in L^1$ gilt:*

(i) Linearität: $E[aX + Y \,|\, \mathscr{F}] = aE[X \,|\, \mathscr{F}] + E[Y \,|\, \mathscr{F}]$

(ii) Monotonie: $X \leq Y \Rightarrow E[X \,|\, \mathscr{F}] \leq E[Y \,|\, \mathscr{F}]$

(iii) Ist X $\mathscr{F}$-messbar und $X, Y, X \cdot Y \in L^1$, so gilt daass

$$E[XY \,|\, \mathscr{F}] = XE[Y \,|\, \mathscr{F}].$$

(iv) Ist $\mathscr{G} \subseteq \mathscr{F}$ eine σ-Algebra, so ist

$$E[E[X \,|\, \mathscr{F}] \,|\, \mathscr{G}] = E[X \,|\, \mathscr{G}].$$

(v) Ist X unabhängig von $\mathscr{F}$, so ist

$$E[X \,|\, \mathscr{F}] = E[X].$$

Dieser Satz zeigt, dass der bedingte Erwartungswert *linear* und *monoton* ist.

Eigenschaft (iv) wird auch *Turm-Eigenschaft* genannt: Bei iterierten bedingten Erwartungswerten kann man sich immer auf die kleinere σ-Algebra zurückziehen.

Besonders nützlich sind auch die Regeln (iii) und (v): Messbare ZV kann man rausziehen und unabhängige Information wegfallen lassen.

Beweis. (i) Wir wählen $F \in \mathscr{F}$. Dann ist

$$\begin{aligned}\int_F E[aX+Y \,|\, \mathscr{F}]dP &= \int_F aX + YdP = a\int_F XdP + \int_F YdP \\ &= a\int_F E[X \,|\, \mathscr{F}]\,dP + \int_F E[Y \,|\, \mathscr{F}]\,dP \\ &= \int_F (aE[X \,|\, \mathscr{F}] + E[Y \,|\, \mathscr{F}])\,dP.\end{aligned}$$

(ii) Wir zeigen $X \geq 0 \Rightarrow E[X \,|\, \mathscr{F}] \geq 0$. (ii) folgt dann mit (i).

Zunächst ist $X\mathbb{1}_{\{E[X \,|\, \mathscr{F}]<0\}} \geq 0$, also auch der Erwartungswert hiervon.

Wäre $P(E[X \,|\, \mathscr{F}] < 0) > 0$, so wäre

$$\int_{\{E[X \,|\, \mathscr{F}]<0\}} E[X \,|\, \mathscr{F}]dP = \int_{\{E[X \,|\, \mathscr{F}]<0\}} XdP < 0,$$

ein Widerspruch.

(iii) Für $F \in \mathscr{F}$ gilt

$$\int_F E[XY \,|\, \mathscr{F}]\,dP = \int_F XY\,dP.$$

Wir wählen eine Folge einfacher Funktionen (X^n), so dass $X^n \uparrow X$ und betrachten $Y \geq 0$.

Dann ist $|X^n Y| = |X^n|Y \leq |X|Y = |XY|$ und ebenso $X^n \cdot E[Y \,|\, \mathscr{F}] \leq |X| \cdot E[Y \,|\, \mathscr{F}]$ mit majorisierter Konvergenz folgt

$$\begin{aligned}\int_F X^n YdP &= \sum_{i=1}^N x^i \int_{F\cap A^i} YdP = \sum_{i=1}^N x^i \int_{F\cap A^i} E[Y \,|\, \mathscr{F}]dP \\ &= \int_F X^n E[Y \,|\, \mathscr{F}]dP \to \int_F XE[Y \,|\, \mathscr{F}]dP.\end{aligned}$$

Ein symmetrisches Argument liefert die Aussage für $Y \leq 0$ und die Behauptung folgt, da

$$\int_F XYdP = \int_F XY^+dP + \int_F XY^-dP.$$

(iv) Wir betrachten $G \in \mathscr{G}$ und erhalten, da $G \in \mathscr{F}$,

$$\int_G E[X \,|\, \mathscr{G}]dP = \int_G XdP = \int_G E[X \,|\, \mathscr{F}]dP = \int_G E[E[X \,|\, \mathscr{F}] \,|\, \mathscr{G}]dP.$$

(v) Ist X von $\mathscr{F}$ unabhängig, so folgt für $F \in \mathscr{F}$

$$\int_F E[X \,|\, \mathscr{F}]dP = \int 1_F XdP = E[1_F]E[X] = \int_F E[X]dP$$

und die Behauptung folgt. □

Im obigen Beweis ist die Messbarkeit (bzgl. $\mathscr{F}$) meist offensichtlich, so dass wir die jeweilige Überprüfung nicht immer explizit aufführen.

Ganz analog zum Erwartungswert erhalten wir aus der Linearität des bedingten Erwartungswertes die Gültigkeit der Jensenschen Ungleichung[40].

[40] siehe Satz 8.

Satz 166. *Sei $\varphi : \mathbb{R} \to \mathbb{R}$ konvex und $X \in L^1$. Dann gilt*

$$E[\varphi(X) \,|\, \mathscr{F}] \geq \varphi(E[X \,|\, \mathscr{F}]).$$

Für den folgenden Satz benötigen wir dass für eine (P-f.s.) monoton wachsende Folge von Zufallsvariablen ein vernünftiger stochastischer Grenzwert existiert.

Ist zB. $\Omega = [0,1]$ mit dem Lesbesgue-Maß, und $\Phi = \{\mathbb{1}_x : x \in [0,1]\}$, so ist $\varphi = 0$ P-f.s., aber $\sup_{\varphi \in \Phi} \varphi = 1$. Es benötigt also ein Konzept für ein Supremum was für Zufallsvariablen geeignet ist.

Satz 167. *Sei Φ eine beliebige Menge von reellwertigen Zufallsvariablen. Dann gibt es eine Zufallsvariable φ^*, so dass*

(i) $\varphi^* \geq \varphi$ *P-f.s. für alle $\varphi \in \Phi$,*
(ii) $\varphi^* \leq \psi$ *für jede Zufallsvariable ψ mit $\psi \geq \varphi$ P-f.s. für alle $\varphi \in \Phi$.*

Diese Zufallsvariable φ^* nennen wir das **essentielle Supremum** und schreiben

$$\varphi^* := \operatorname{ess\,sup} \Phi = \operatorname*{ess\,sup}_{\varphi \in \Phi} \varphi.$$

Beweis. Wir können ohne Einschränkung annehmen, dass alle $\varphi \in \Phi$ Werte in $[0,1]$ annehmen – ansonsten betrachten wir $f(\varphi)$ mit einer streng monoton wachsenden Funktion $f : \mathbb{R} \to [0,1]$.

Ist die Teilmenge $\Psi \subset \Phi$ abzählbar, so ist $\varphi_\Psi := \sup_{\varphi \in \Psi} \varphi$ messbar. Die Erwartungswerte der Zufallsvariablen $\varphi \in \Phi$ werden von φ^* maximiert. Aus diesem Grund betrachten wir

$$c := \sup\{E[\varphi_\Psi] : \Psi \subset \Phi \text{ und } \Psi \text{ ist abzählbar }\}.$$

Wir wählen eine Folge Ψ_n, so dass $E[\varphi_{\Psi_n}] \to c$ und setzen $\Psi^* := \cup \Psi_n$. Dann ist Ψ^* wieder abzählbar und $E[\varphi_{\Psi^*}] = c$.

Wir zeigen, dass $\varphi^* := \varphi_{\Psi^*}$ die beiden Bedingungen erfüllt: Angenommen, (i) gelte nicht. Dann gibt es ein $\varphi \in \Phi$, so dass $P(\varphi > \varphi^*) > 0$ ist. Dann wäre aber

$$E[\varphi_{\Psi^* \cup \varphi}] > c,$$

ein Widerspruch.

Ist $\psi \geq \phi$ für all $\varphi \in \Phi$, so auch für alle $\varphi \in \Psi^* \subset \Phi$, so dass auch (ii) gilt. $\square$

Wir sehen aus diesem Beweis auch, dass man eine Approximation finden kann, so lange Φ folgende Eigenschaft erfüllt: für $\varphi_1, \varphi_2 \in \Phi$ gibt es ein $\psi \in \Phi$, so dass $\psi \geq \max\{\varphi_1, \varphi_2\}$. Dann finden wir (φ_n), so dass $\varphi_n \to \operatorname{ess\,sup} \Phi$.

Satz 168. *Seien $X^1, X^2, \ldots \in L^1$ und entweder*

(i) $X \in L^1$ *und* $X^n \uparrow X$ *f.s. oder*

(ii) $X_n \to X$ *f.s. und* $\exists Y \in L^1$*, so dass* $|X^n| \leq |Y|\ \forall n$.

Dann gilt

$$E[X_n \mid \mathscr{F}] \to E[X \mid \mathscr{F}] \ f.s. \text{ und in } L^1.$$

Beweis. Für die L^1-Konvergenz hat man mit der Jensenschen Ungleichung

$$\begin{aligned} E\Big[\big|E[X_n \mid \mathscr{F}] - E[X \mid \mathscr{F}]\big| \Big] &= E\Big[\big|E[X_n - X \mid \mathscr{F}]\big| \Big] \\ &\leq E[|X_n - X|] \longrightarrow 0. \end{aligned}$$

Für die fast sichere Konvergenz unterscheiden wir die beiden Fälle:

(i) Zunächst liefert der Satz der monotonen Konvergenz

$$\begin{aligned} \int_F X dP &= \lim_{n\to\infty} \int_F X^n dP \\ &= \lim_{n\to\infty} \int_F E[X^n \mid \mathscr{F}]\, dP \end{aligned}$$

für alle $F \in \mathscr{F}$. Da auch $E[X^n \mid \mathscr{F}]$ monoton wächst, ist weiterhin

$$\lim_{n\to\infty} \int_F E[X^n \mid \mathscr{F}]\, dP = \int_F \lim_{n\to\infty} E[X^n \mid \mathscr{F}]\, dP.$$

Der Grenzwert dieser monoton wachsenden Folge ist das essentielle Supremum,

$$E[X^n \mid \mathscr{F}] \leq E[X^{n+1} \mid \mathscr{F}] \to \operatorname*{ess\,sup}_n E[X^n \mid \mathscr{F}] =: Y,$$

und somit folgt $Y = \lim_{n\to\infty} E[X^n \mid \mathscr{F}] = E[X \mid \mathscr{F}]$.

(ii) Wir setzen

$$\begin{aligned} Y^n &= \operatorname*{ess\,sup}_{k\geq n} X^k \downarrow X \ f.s. \text{ und} \\ Z^n &= \operatorname*{ess\,inf}_{k\geq n} X^k \uparrow X \ f.s. \end{aligned}$$

Dann ist

$$-Y \leq Z^n \leq X^n \leq Y^n \leq Y, \text{ also } Y^n, Z^n \in L^n.$$

Mit (i) erhalten wir

$$E[X \,|\, \mathscr{G}] = \lim_{n\to\infty} E[Z^n \,|\, \mathscr{F}] \le \lim_{n\to\infty} E[X^n \,|\, \mathscr{F}] \le \lim_{n\to\infty} E[Y^n \,|\, \mathscr{F}] = E[X \,|\, \mathscr{F}]$$

fast sicher und die Behauptung folgt. □

Bemerkung 169. Die bedingte Erwartung kann man in L^2 auch als Projektion auffassen: In dem Hilbertraum $H = L^2$ gilt für einen abgeschlossenen linearen Teilraum M, dass jedes $X \in H$ zerlegt werden kann in

$$X = Y + Z$$

mit $Y \in M$ und $Z \perp M$. Wir setzen $M = \{X' : X' = Y \; f.s., Y \text{ ist } \mathscr{F}\text{-messbar}\}$.

Dann ist $X - E[X \,|\, \mathscr{F}] \perp M$, also

$$0 = \langle 1_F, X - E[X \,|\, \mathscr{F}] \rangle = \int_F (X - E[X \,|\, \mathscr{F}]) dP$$

und Y ist in der Tat die bedingte Erwartung von X gegeben L.

Eine besondere Rolle kommt dem Fall zu, wo $\mathscr{F}$ von einer Zufallsvariable erzeugt wird.

Wir hatten bereits gezeigt, dass eine Zufallsvariable Y genau dann $\sigma(Z)$ messbar ist, falls es eine Funktion f gibt, so dass $Z = f(z)$.

Definition 170. Ist $\mathscr{F} = \sigma(Z)$, so heißt die Funktion

$$f(z) := E[X \,|\, Z = z]$$

gegeben durch $E[X \,|\, Z] = f(Z)$ *Faktorisierung* von X gegeben $\mathscr{F}$.

Beispiel 171. Seien $X_1, X_2, \dots$ iid mit $\mu = E[X_1] < \infty$. Wir setzen $S_n = \sum_{i=1}^n X_i$. Dann ist

$$E[S_n \,|\, X_i] = E[X_1 \,|\, X_1] + E[X_2 + \cdots + X_n \,|\, X_1] = X_1 + (n-1) \cdot \mu.$$

Und:

$$E[X_1 \,|\, S_n] \overset{(*)}{=} \frac{1}{n} \sum_{i=1}^n E[X_i \,|\, S_n] = \frac{1}{n} E[S_n \,|\, S_n] = \frac{1}{n} S_n.$$

$(*)$: X_i sind iid

Aufgabe 4. Bedingte Wahrscheinlichkeiten erfüllen bis auf Nullmengen die Axiome von Wahrscheinlichkeitsmaßen.

Beispiel 172 (Die bedingte Dichte)**.** Haben die Zufallsvariablen X, Y die gemeinsame Dichte $f(x,y)$, so ist die bedingte Dichte gegeben durch:

$$f(x \,|\, y) = \frac{f(x,y)}{f(y)} \text{ falls } f(y) > 0$$

Die bedingte Dichte beschreibt die bedingte Verteilung:

$$P(X \leq x \mid Y = y) = \int_{-\infty}^{x} f(z \mid y) dz.$$

Beweis. Für alle $A, B \in \mathcal{B}(\mathbb{R})$ gilt

$$P(X \in A, Y \in B) = \int_{B \cap f(y) > 0} \int_A f(x,y) dx dy = \int_{B \cap f(y) > 0} \underbrace{\int_A \frac{f(x,y)}{f(y)} dx}_{=: g(y)} f(y) dy$$

Wir erhalten also

$$\int_{Y \in B} E[\mathbb{1}_{\{X \in A\}} \mid Y] dP = \int_{Y \in B} g(y) dP$$

und somit ist $g(y) = g_A(y)$ die Faktorisierung von $P(X \in A | Y)$. Daraus folgt natürlich, dass

$$P(X \leq x | y) = \int_{-\infty}^{x} \frac{f(x,y)}{f(y)} dx$$

so lange $f(y) > 0$ und somit die Behauptung. □

Beispiel 173. Sind X, Y gemeinsam standard-normalverteilt mit Korrelation ϱ, so ist

$$E[X \mid Y] = \varrho Y.$$

Die beweist man wie folgt:

$$f(x,y) = e^{-\frac{1}{2} \frac{(x^2 - \varrho 2xy + y^2)}{1 - \varrho^2}} \frac{1}{2\pi}$$
$$f(y) = e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}}$$

$$\Rightarrow \qquad f(x \mid y) = \frac{1}{\sqrt{2\pi}} \exp\Big(\underbrace{-\frac{1}{2} \frac{x^2 - 2\varrho xy + y^2}{1 - \varrho^2} - \frac{-y^2(1 - \varrho^2)}{1 - \varrho^2}}_{= -\frac{1}{2} \frac{(x - \varrho y)^2}{(1 - \varrho^2)}} \Big).$$

Man beachte, dass $X \mid Y$ wieder normalverteilt ist, allerdings mit verringerter Varianz: $1 - \varrho^2 < 1$ (falls $\varrho \neq 0$).

Dieser einfache Zusammenhang ist die Basis des **Kalman-Filters**.

Für eine tolle Darstellung und vielfältige weitere Verweise siehe `https://de.wikipedia.org/wiki/Kalman-Filter`

Ein Algorithmus zur Berechnung von Größen, die unter gestörter Beobachtung gemacht werden. Sie wurde unter anderem bei der Apollo-Mission zum Mond eingesetzt (und bei allen GPS-Systemen).

Aufgabe 5. Die Faktorisierung löst die Gleichung $(E[X \mid Y] = f(Y))$

$$\int_B f(y) dP_y(y) = \int_{Y \in B} X dP \qquad \text{für alle } B \in \mathscr{B}(\mathbb{R}) \qquad (174)$$

Lemma 175. *Sei X von Z unabhängig. Dann folgt für jede messbare Funktion $T : \mathbb{R}^2 \to \mathbb{R}$ s. d. $E[|T(X,y)|] < \infty$ für alle $y \in \mathbb{R}$,*

$$E[T(X,Z) \mid Z = z] = E[T(X,z)].$$

Beweis. Für alle $A \in \mathscr{B}(\mathbb{R})$ ist

$$\begin{aligned} \int_A E[T(X,Z) \mid Z = z]\, dP^Z(z) &\overset{(174)}{=} \int_{Z \in A} T(X,Z) dP \\ &\overset{(*)}{=} \int_A \underbrace{\int_{\mathbb{R}} T(X,z) dP^X(x)}_{=E[T(X,z)]} dP^Z(z) \end{aligned}$$

$(*)$: X, Z unabhängig

und die Behauptung folgt. □

Stochastische Prozesse und Martingale

Nun kommen wir zu dem wichtigen Begriff |textbfMartingal.

Definition 176. Eine *Filtration* $\mathbb{F} = (\mathscr{F}_n)_{n \geq 1}$ ist eine Folge von wachsenden Sub-σ-Algebren.

Wachsend heißt hierbei schlicht, dass $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$, es geht also keine Information verloren. Außerdem ist stets $\mathscr{F}_n \subset \mathscr{A}$.

Definition 177. Eine Folge $X = (X_n)_{n \geq 1}$ von Zufallsvariablen heißt *stochastischer Prozess* (in diskreter Zeit). Zu einem stochastischen Prozess $X = (X_n)$ heißt $\mathbb{F}^X = (\mathscr{F}_n^X)$ gegeben durch

$$\mathscr{F}_n^X = \sigma(X_1, \ldots, X_n)$$

die *natürliche* Filtration von X.

Im Folgenden betrachten wir stets eine feste Filtration $\mathbb{F}$, die nicht die natürliche Filtration von X sein muss.

Definition 178. Einen stochastischer Prozess X nennt man *vorhersehbar* (bezüglich $\mathbb{F}$), wenn X_n $\mathscr{F}_{n-1}$-messbar ist, für alle $n \geq 1$. X nennt man *adaptiert* (an $\mathbb{F}$), wenn X_n $\mathscr{F}_n$-messbar ist für alle $n \geq 1$.

Definition 179. Ein adaptierter stochastischer Prozess X heißt

Martingal,	falls $E[X_{n+1} \mid \mathscr{F}_n] = X_n$,
Submartingal,	falls $E[X_{n+1} \mid \mathscr{F}_n] \geq X_n$,
Supermartingal,	falls $E[X_{n+1} \mid \mathscr{F}_n] \leq X_n$,

Beispiel 180 (Fortsetzung)**.** Sind (X_n) iid und

$$X_n = \sum_{i=1}^{n} Y_i,$$

so ist X ein Martingal (Sub-/Super-), falls $\mu = 0$ (> 0, < 0).

Lemma 181. *Die Bedingungen in Definition 179 sind jeweils äquivalent zu*

$$E[X_m \mid \mathscr{F}_n] = X_n$$

für alle $m \geq n$ (bzw. $\geq$, $\leq$ für Sub-/Supermartingal).

Beispiel 182 (Das Black-Scholes-Modell)**.** Wir betrachten eine Tages-Rendite

$$Z_n := \exp(m + \sigma \cdot \xi_n),$$

wobei (ξ_n) iid $\mathcal{N}(0,1)$ sind. Das Black-Scholes-Modell ist gegeben durch

$$S_n = S_0 \prod_{i=1}^{n} Z_n = S_0 \exp\left(m \cdot n + \sigma \sum_{i=1}^{n} \xi_i\right).$$

Wir erhalten

$$E[S_{n+1}] \mid \mathscr{F}_n] = e^m E[e^{\sigma \xi_1}] \cdot S_n.$$

Da $\xi \sim \mathcal{N}(0,1)$ folgt

$$E[e^{\sigma\xi}] = e^{\frac{\sigma^2}{2}}.$$

S ist also genau dann ein Martingal, falls $m = -\frac{\sigma^2}{2}$.

Satz 183 (Doob-Meyer-Zerlegung)**.** *Jeder adaptierte Prozess X besitzt eine Zerlegung*

$$X = A + M,$$

wobei M ein Martingal ist und A vorhersehbar ist. Ist $A_1 = 0$, so ist die Zerlegung f.s. eindeutig.

Beweis. Wir definieren

$$A_n = \sum_{i=2}^{n} E[X_i - X_{i-1} \,|\, \mathscr{F}_{i-1}]. \qquad A_1 = 0. \tag{184}$$

Dann ist A vorhersehbar und $M := X - A$ ein Martingal, denn

$$E[M_n - M_{n-1} \,|\, \mathscr{F}_{n-1}] = E[X_n - X_{n-1} \,|\, \mathscr{F}_{n-1}] - \underbrace{(A_n - A_{n-1})}_{=E[X_n - X_{n-1} \,|\, \mathscr{F}_{n-1}]} = 0.$$

Zur Eindeutigkeit: Gilt $X = M + A$ mit einem Martingal M, so ist

$$\begin{aligned} E[X_n - X_{n-1} \,|\, \mathscr{F}_{n-1}] &= E[M_n - M_{n-1} \,|\, \mathscr{F}_{n-1}] + (A_n - A_{n-1}) \\ &= M_{n-1} - M_{n-1} + (A_n - A_{n-1}) = A_n - A_{n-1}, \end{aligned}$$

also gilt (184) fast sicher. □

Definition 185. Eine *zufällige Zeit* ist eine messbare Abbildung

$$\tau : \Omega \to \{1, 2, \ldots, \infty\}.$$

Sie heißt $(\mathbb{F})$-Stoppzeit, falls

$$\{\tau = n\} \in \mathscr{F}_n \qquad \forall\, n \geq 1.$$

Bemerkung 186. τ ist genau dann eine Stoppzeit, wenn

$$\{\tau \leq n\} \in \mathscr{F}_n \qquad \forall\, n \geq 1.$$

Definition 187. Für eine Stoppzeit τ definieren wir durch

$$\mathscr{F}_\tau := \{F \in \mathscr{A} : F \cap \{\tau = n\} \in \mathscr{F}_n \quad \forall n \geq 0\}$$

die σ-Algebra der τ-Vergangenheit.

Aufgabe 6. Gilt $\tau \leq \sigma$ für zwei Stoppzeiten τ, so ist $\mathscr{F}_\tau \subseteq \mathscr{F}_\sigma$.

Lemma 188. *Ist X adaptiert, so ist X_τ $\mathscr{F}_\tau$-messbar. Ebenso ist τ $\mathscr{F}_\tau$-messbar.*

Beweis. Zunächst ist $\{\tau = m\} \in \mathscr{F}_\tau$, da

$$\{\tau = m\} \cap \{\tau = n\} = \begin{cases} \{\tau = m\} & \text{falls } m = n, \\ \varnothing & \text{sonst.} \end{cases}$$

In beiden Fällen ist die Menge $\mathscr{F}_\tau$-messbar.

Ebenso ist

$$\{X_\tau \in A\} \cap \{\tau = n\} = \{X_n \in A\} \cap \{\tau = n\} \in \mathscr{F}_n, \quad \forall A \in \mathscr{B}(\mathbb{R}),$$

also ist X_τ $\mathscr{F}_\tau$-messbar. □

Für einen Prozess X und eine Stoppzeit τ bezeichnen wir durch

$$X^\tau = (X_{\tau \wedge n})_{n \geq 1}$$

den an τ gestoppten Prozess.

Satz 189. *Ist X ein (Sub-/Super-)Martingal, so ist X^τ ein (Sub-/Super-) Martingal.*

Beweis. Zunächst ist

$$X_{n\wedge\tau} = \sum_{i=1}^{n-1} \underbrace{X_i}_{\in\mathscr{F}_n} \underbrace{\mathbb{1}_{\{\tau=i\}}}_{\in\mathscr{F}_n} + \underbrace{X_n}_{\in\mathscr{F}_n} \underbrace{\mathbb{1}_{\{\tau\geq n\}}}_{\in\mathscr{F}_n}$$

Damit ist $X_{n\wedge\tau}$ $\mathscr{F}_n$-messbar, also X^τ adaptiert. Weiterhin ist

$$(X_{(n+1)\wedge\tau} - X_{n\wedge\tau}) = \mathbb{1}_{\{\tau>n\}}(X_{n+1} - X_n), \quad \text{also}$$

$$E[X^\tau_{n+1} - X^\tau_n \mid \mathscr{F}_n] = \mathbb{1}_{\{\tau>n\}} E[X_{n+1} - X_n \mid \mathscr{F}_n] = 0 \qquad (\geq 0 \mid \leq 0),$$

falls X ein Martingal (Sub-/Supermartingal), und die Behauptung folgt. □

Beispiel 190. Typische Beispiele von Stoppzeiten sind Ersteintrittszeiten.

Bisher haben wir stets (!) endliche Stoppzeiten betrachtet. In vielen (aber nicht allen) Fällen kann das abgeschwächt werden.

Beispiel 191 (Symmetrische Irrfahrt)**.** Seien $X_n = \sum_{i=1}^n \xi_i$, $\xi_i \in \{-1, 1\}$ iid mit $P(\xi_i = 1) = \frac{1}{2}$. Dann ist

$$T_k := \inf\{k \geq 1 : X_n = k\}$$

eine Stoppzeit! Außerdem ist $P(T_k < \infty) = \mathbb{1}$. Aber

$$\mathbb{1}_{\{T_k<\infty\}} S_{T_k} = k \mathbb{1}_{\{T_k<\infty\}},$$

also $E[X_{T_k}] = k$ und somit kann X^{T_k} kein Martingal sein.

Für den folgenden Satz bezeichnen wir wieder

$$S_n = \sum_{i=1}^{n} X_i.$$

Satz 192 (Wald). *Seien $(X_n)_{n\geq 1}$ identisch verteilt, adaptiert mit $E[|X_1|] < \infty$, und X_{n+1} sei von $\mathscr{F}_n$ unabhängig für alle $n \geq 1$. Ist τ Stoppzeit und $E[\tau] < \infty$, so ist*

$$E[S_\tau] = E[\tau] \cdot E[X_1].$$

Gilt außerdem $E[X_n^2] < \infty$, $E[X_n] = 0$, so folgt

$$E[S_\tau^2] = E[\tau] \cdot E[X_1^2].$$

Beweis. Für ein beschränktes τ könnte man direkt S^τ betrachten und wäre fertig. Die allgemeinere Voraussetzung macht ein anderes Vorgehen nötig. Zunächst folgt aus $E[\tau] < \infty$, dss $P(\tau = \infty) = 0$. Weiterhin ist

$$\begin{aligned} E[|S_\tau|] &= \sum_{n\geq 1} E[|S_\tau| \mathbb{1}_{\{\tau=n\}}] \leq \sum_{n\geq 1} \sum_{i=1}^{n} E[|X_i| \mathbb{1}_{\{\tau=n\}}] \\ &\overset{(*)}{=} \sum_{i\geq 1} \sum_{n=i}^{\infty} E[|X_i| \mathbb{1}_{\{t=n\}}] = \sum_{i\geq 1} E[|X_i| \mathbb{1}_{\{\tau\geq i\}}] \end{aligned} \tag{193}$$

Die Vertauschung der Summen in (*) ist unter Endlichkeit erlaubt, was wir gleich in (194) sehen werden. Nun ist aber $\mathbb{1}_{\{\tau\geq i\}} = 1 - \mathbb{1}_{\{\tau<i\}} = 1 - \mathbb{1}_{\{\tau<i-1\}}$ $\mathscr{F}_{i-1}$-messbar, also

$$E[|X_i| \mathbb{1}_{\{\tau\geq i\}}] = E[|X_i|] \cdot P(\tau \geq i),$$

und wir erhalten die Integrierbarkeit von S_τ

$$E[|S_\tau|] \leq \sum_{i\geq 1} P(\tau \geq i) \underbrace{E[|X_i|]}_{= E[|X_1|]} = E[\tau] \cdot E[|X_1|] < \infty. \tag{194}$$

Wieder einmal nutzen wir die praktische Darstellung, dass $E[\tau] = \sum P(\tau \geq n)$ falls τ endliche Stoppzeit ist.

Dieselben Überlegungen mit S_τ anstelle von $|S_\tau|$ liefern den ersten Teil.

Für den zweiten Teil betrachten wir

$$Y_n = X_n \mathbb{1}_{\{\tau\geq n\}}.$$

Für $m > n$ gilt

$$E[Y_m Y_n] = E[X_n \mathbb{1}_{\{\tau\geq n\}} \mathbb{1}_{\{\tau\geq m\}} \underbrace{E[X_m \mid \mathscr{F}_{m+1}]}_{=0}],$$

so dass die Kovarianzen verschwinden. Wir erhalten

$$\sum_{n\geq 1} E[Y_n^2] = E[X_1^2] \sum_{n\geq 1} P(\tau \geq n) = E[X_1^2] \cdot E[\tau] < \infty. \tag{195}$$

Damit ist

$$\begin{aligned} \|Y_m + \cdots + Y_n\|^2 &= E[(Y_m + \cdots + Y_n)^2] \\ &= E[Y_m^2] + \cdots + E[X_n^2], \end{aligned}$$

und wir erhalten L_2-Konvergenz von $\sum_{i=1}^n Y_i^2 \xrightarrow{L_2} \sum_{i\geq 1} Y_i^2$ aus Konvergenz der Reihe in (195). Da L^2 abgeschlossen ist, ist der Grenzwert $\sum_{i\geq 1} Y_i^2$ in L^2. Außerdem ist wie in Gleichung (193)

$$E[S_\tau^2] = \lim_{n\to\infty} E\Big[\Big(\sum_{i=1}^n Y_i\Big)^2\Big] = \sum_{i=1}^\infty E[Y_i^2] = E[X_1^2] \cdot E[\tau].$$

Aus Teil 1 folgt auch, dass $E[S_\tau] = E[\tau]E[X_1] = 0$. □

Wichtige Ungleichungen

Als erstes beweisen wir die wichtige Maximal-Ungleichung von Kolmogorov. Sie überträgt die Markov-Ungleichung auf eine gleichmäßige Aussage für Submartingale.

Satz 196 (Maximal-Ungleichung)**.** *Ist X ein Submartingal und $\alpha > 0$, so folgt*

$$P\big(\max_{1\leq i\leq n} X_i \geq \alpha\big) \leq \frac{1}{\alpha} E[|X_n|].$$

Beweis: Sei $\tau_2 = n$ und $\tau_1 = \inf\{k \geq 1 : X_k \geq \alpha\}$ mit $\inf \varnothing = n$. Ist $M_k = \max_{1\leq i\leq k} X_i$, so folgt daraus

$$\{M_n \geq \alpha\} \cap \{\tau_1 \leq k\} = \{M_k \geq \alpha\} \in \mathscr{F}_k,$$

also $\{M_n \geq \alpha\} \in \mathscr{F}_{\tau_1}$.

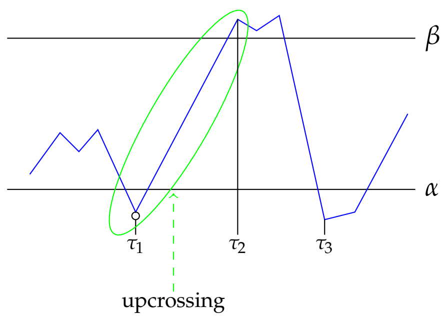
Nun folgt aus $M_n \geq \alpha$, dass $X_{\tau_1} \geq \alpha$, also

$$\begin{aligned} \alpha P(M_n \geq \alpha) &\leq \int_{\{M_n\geq\alpha\}} X_{\tau_1} dP \overset{X \text{ submart.}}{\leq} \int_{\{M_n\geq\alpha\}} X_n dP \\ &\leq \int_{\{M_n\geq\alpha\}} X_n^+ dP \leq E[X_n^+] \leq E[|X_n|]. \end{aligned}$$ □

Wir bemerken, dass aus der letzten Gleichung auch folgt, dass

$$P\big(\max_{1\leq i\leq n} X_i \geq \alpha\big) \leq \frac{1}{\alpha} E[|X_n|\, \mathbb{1}_{\{\max_{1\leq i\leq n} X_i \geq \alpha\}}]. \tag{197}$$

Die zweite wichtige Ungleichung benutzt Upcrossings: Geben wir eine untere Schranke α und eine obere Schranke β vor, so ist ein Upcrossing ein Durchgang des Prozesses aus dem Bereich unterhalb von α in den Bereich oberhalb von β. Dies kann man natürlich gut über Stoppzeiten beschreiben.



Wir betrachten $\alpha < \beta$ und definieren $\tau_1 := \inf\{i \geq 1 : X_i \leq \alpha\}$ mit $\inf \varnothing = n$ und

$$\tau_k := \begin{cases} \inf\{k > \tau_{k-1} : X_i \geq \beta\}, & k \text{ gerade} \\ \inf\{k > \tau_{k-1} : X_i \leq \alpha\}, & k \text{ ungerade} \end{cases}$$

Wir definieren die Anzahl der Upcrossings in dem Intervall $[0, n]$ durch:

$$U^n_{[\alpha,\beta]} \text{ als das größte } k, \text{ so dass } X_{\tau_{2k-1}} \leq \alpha < \beta \leq X_{\tau_{2k}}$$

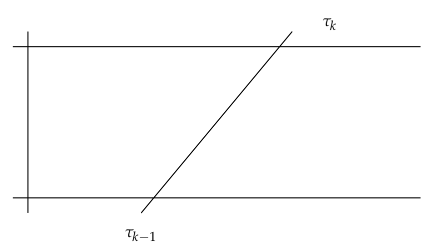
Satz 198 (Upcrossing Inequality)**.** *Sei X ein Submartingal. Dann ist*

$$E[U^n_{[\alpha,\beta]}] \leq \frac{E[|X_n|] + |\alpha|}{\beta - \alpha}.$$

Beweis. Wir setzen $Y_k = \max(X_k - \alpha, 0) = (X_k - \alpha)^+$ und $\theta = \beta - \alpha$.

Dann ist die Fragestellung äquivalent zu den $U[0, \theta]$-Upcrossings von $Y_1, \ldots, Y_n$. Wegen Jensen's Ungleichung ist auch Y ein Submartingal.

Wir betrachten ein gerades $k < n$ (wie im Bild rechts). Dann ist



$$\{\tau_k = j\} = \bigcup_{i=1}^{j-1} \underbrace{\{\tau_{k-1} = i, Y_{i+1} < \theta, \ldots, Y_{j-1} < \theta, Y_j \geq \theta\}}_{\in \mathscr{F}_j} \in \mathscr{F}_j$$

und $\{\tau_k = n\} = \{\tau_k \leq n-1\}^c \in \mathscr{F}_n$, also ist τ_k Stoppzeit. Ein ähnliches Argument zeigt dies auch für k ungerade.

Da die τ_k strikt wachsen, ist $\tau_n = n$.

$$Y_n = Y_{\tau_m} \geq Y_{\tau_n} - Y_{\tau_1} = \sum_{k=2}^{n} (Y_{\tau_k} - Y_{\tau_{k-1}}).$$

Diese Summe teilen wir auf in

$$A = \sum_{k=2,\ k \text{ gerade}}^{n} (Y_{\tau_k} - Y_{\tau_{k-1}})$$

und $B = Y_{\tau_n} - Y_{\tau_1} - A$.

Betrachten wir den Ausdruck B, also ungerades k. Y ist ein Submartingal, so dass $E[Y_{\tau_k}] \geq E[Y_{\tau_{k-1}}]$, also $E[B] \geq 0$. Weiterhin ist $Y_n \geq A + B$. Mit $E[B] \geq 0$ folgt $E[Y_n] \geq E[A]$.

Nun schätzen wir $E[A]$ noch ab. Für gerades k mit $\tau_k < n$ ist $Y_{\tau_n} - Y_{\tau_1} \geq \theta$, also

$$A \geq U_{[\alpha,\beta]} \cdot \theta.$$

Hieraus folgt natürlich $E[Y_n] \geq \theta \cdot E[U_{[\alpha,\beta]}]$. Formuliert in X erhalten wir

$$(\beta - \alpha)E[U_{[\alpha,\beta]}] \leq E[(X_n - \alpha)\mathbb{1}_{\{X_n - \alpha \geq 0\}}] \leq E[|X_n|] + |\alpha|. \tag{199}$$

Die Gleichung (199) ermöglicht noch eine etwas schärfere Abschätzung, falls nötig:

$$(\beta - \alpha)E[U_{[\alpha,\beta]}] \leq E[(X_n - \alpha)^+].$$

□

Wir erhalten ein erstes Konvergenzresultat.

Satz 200. *Sei (X_n) ein Submartingal und $K = \sup_n E[|X_n|] < \infty$. Dann gibt es eine Zufallsvariable X mit $E[|X|] \leq K$, so dass*

$$X_n \xrightarrow{f.s.} X.$$

Beweis. Wir fixieren α, β und notieren mit $U^n = U^n_{[\alpha,\beta]}$ die Upcrossings von $X_1, \ldots, X_n$. Nach dem Upcrossing-Lemma ist

$$E[U^n] \leq \frac{E[|X_n|] + |\alpha|}{\beta - \alpha} \leq \frac{K + |\alpha|}{\beta - \alpha}$$

Nun ist U^n wachsend und hat eine integrierbare Majorante also folgt mit dem Satz von der majorisierten Konvergenz, dass $\sup U^n \in L_1$, also f. s. endlich ist.

Wir setzen $X^* = \limsup X_n$ und $X_* = \liminf X_n$. Ist $X_* < \alpha < \beta < X^*$, dann gilt $U^{=}U^n_{[\alpha,\beta]} \to \infty$, also $P(X_* < \alpha < \beta < X^*) = 0$. Nun ist

$$\{X_* < X^*\} = \bigcup \{X_* < \alpha < \beta < X^*\},$$

woebei wir über alle $\alpha < \beta \in \mathbb{Q}$ summieren. Da die rechte Seite die Wahrscheinlichkeit 0 hat, so auch die linke und $X_n \to X^* = X_*$ f. s.

Bisher ist $X^* \in \mathbb{R} \cup \{-\infty, \infty\}$. Nach Fatou ist

$$E[|X|] \leq \liminf E[|X_n|] \leq K, \text{ also } P(X \in \mathbb{R}) = 1.$$ □

Satz 201 (Doob's L^p-Ungleichung). *Sei X ein Martingal oder ein positives Supermartingal und $p > 1$. Dann gilt*

$$E[\sup_{m \le n} |X_m|^p] \le \Big(\frac{p}{p-1}\Big)^p E[|X_n|^p].$$

Bemerkung 202. Eine Schranke nach unten ist leicht zu finden:

$$E[|X_n|^p] \le E[\sup_{m \le n} |X_m|^p].$$

Beweis. Die Idee ist, die Maximalungleichung anzuwenden. Man nutzt

$$(y \wedge K)^p = \int_0^{y \wedge K} p\, z^{p-1} dz = \int_0^K \mathbb{1}_{\{z<y\}} p\, z^{p-1} dz.$$

Somit ist

$$\begin{aligned}
E\Big[\sup_{m \le n}(|X_m| \wedge K)^p\Big] &= E[(\sup_{m \le n} |X_m| \wedge K)^p] && x^p \text{ monoton auf } \mathbb{R}_{\ge 0}\\
&= E\Big[\int_0^K \mathbb{1}_{\{z<\sup |X_m|\}} p z^{p-1} dz\Big]\\
&\overset{\text{Fubini}}{=} \int_0^K p z^{p-1} P(\sup |X_m| > z) dz\\
&\overset{\text{Max.Ungl., (197)}}{\le} \int_0^K p z^{p-2} E[|X_n| \mathbb{1}_{\{\sup_{m \le n} |X_m| \ge z\}}] dz && |X| \text{ Submartingal nach Jensen}\\
&\overset{\text{Fubini}}{=} pE[|X_n| \int_0^{\sup |X_m| \wedge K} z^{p-2} dz]\\
&= \frac{p}{p-1} E[|X_n| (\sup |X_m| \wedge K)^{p-1}]
\end{aligned}$$

Nun verwenden wir die Hölder-Ungleichung $\|x \cdot y\| = \|x\|_p \|x\|_q$ mit $\frac{1}{p} + \frac{1}{q} = 1$, also $q = \frac{p}{p-1}$

$$\le \frac{p}{p-1} \|X_n\|_p E[(\sup |X_m| \wedge K)^p]^{\frac{p-1}{p}}.$$

Wir potenzieren beide Seiten mit p und teilen durch $E[(\sup |X_m| \wedge K)^p]^{p-1} \Rightarrow$

$$E[\sup_{n \le m}(|X_m|)^p] \underset{\substack{\uparrow\\ \text{monot.K.}}}{=} \lim_{K \to \infty} E[\sup(|X_m| \wedge K)^p] \le \Big(\frac{p}{p-1}\Big)^p E[|X_n|^p]$$

Man sieht direkt, dass die Argumentation auch für positive Submartingale funktioniert. □

Doob's Martingalkonvergenzsatz

Wir beginnen mit zwei kurzen Beobachtungen zu bedingten Erwartungswerten.

Lemma 203. *Gilt $X_n \xrightarrow{L^1} X$ und sind $X_n, X \in L^1$, so folgt für jede Sub-σ-Algebra $\mathscr{F}$, dass*

$$E[X_n \mid \mathscr{F}] \xrightarrow{\ L^1\ } E[X \mid \mathscr{F}]$$

Beweis. Die bedingten Erwartungswerte existieren, da $X_n, X \in L^1$ nach Voraussetzung. Wir erhalten

$$E[|E[X_n \mid \mathscr{F}] - E[X \mid \mathscr{F}]|] = E[|E[X_n - X \mid \mathscr{F}]|] \leq E[|X_n - X|] \to 0. \quad \square$$

Lemma 204. *Ist $X \in L^1$, so ist $(E[X \mid \mathscr{F}_n])_{n\in\mathbb{N}}$ gleichgradig integrierbar.*

Beweis. Wir verwenden Bedingung (iii) aus Lemma 24: Da $X \in L^1$, gilt mit der Jensenschen Ungleichung

$$\begin{aligned} E\big[(E[X \mid \mathscr{F}_n] - K)_+\big] &= E\big[(E[X - K \mid \mathscr{F}_n])_+\big] \\ &\leq E\big[E[(X-K)_+ \mid \mathscr{F}_n]\big] \to 0 \end{aligned}$$

mit $K \to \infty$ und die Behauptung folgt. $\square$

Der Kern des nächsten Theorems ist, dass Martingale (bzw. Sub- oder Supermartingale) genau dann abschließbar sind (d.h. es gibt X_∞, so dass $X_n = E[X_\infty|\mathscr{F}_n]$), falls sie gleichgradig integrierbar sind.

Theorem 205. *Sei $\mathscr{F}_\infty = \sigma(\bigcup_{n\geq 1} \mathscr{F}_n)$ und $X = (X_{n\geq 1})$ ein Sub-(Super-)Martingal. Dann sind äquivalent:*

(i) $\{X_n : n \geq 1\}$ *ist gleichgradig integrierbar,*
(ii) *es existiert $X_\infty \in L^1(\mathscr{F}_\infty)$, so dass*

$$X_n \xrightarrow{f.s., L^1} X_\infty,$$

(iii) *es existiert $X_\infty \in L^1(\mathscr{F}_\infty)$, so dass $X_n \to X_\infty$ f.s., $(X_n)_{n\in\mathbb{N}\cup\{\infty\}}$ ein Sub- oder Supermartingal ist und $E[X_n] \to E[X_\infty]$.*

Im Folgenden schreiben wir $X_\infty \in L^0(\mathscr{F}_\infty)$ für eine $\mathscr{F}_\infty$-messbare Zufallsvariable und $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$.

Beweis. Wir beginnen mit (i) $\Rightarrow$ (ii): Da $(X_n)_{n\geq 1}$ gleichgradig integrierbar sind, folgt zunächst $\sup E[|X_n|] < \infty$ (nach Lemma 24). Nach Satz 200 folgt die Existenz von $X_\infty \in L^0(\mathscr{F}_\infty)$, so dass

$$X_n \xrightarrow{f.s.} X_\infty.$$

Durch gleichgradige Integrierbarkeit gilt mit dem Konvergenzsatz von Vitali, Theorem 26, auch, dass $X_n \xrightarrow{L^1} X_\infty$ und $X_\infty \in L^1$.

Nun beweisen wir (ii) $\Rightarrow$ (iii) für ein Submartingal X, die Aussage für ein Supermartingal folgt analog. Da $X_n \xrightarrow{L^1} X_\infty$ und sowohl $X_n \in L^1$ als auch $X_\infty \in L^1$, folgt nach Lemma 203, dass für festes n

$$E[X_m \,|\, \mathscr{F}_n] \xrightarrow{L^1} E[X_\infty \,|\, \mathscr{F}_n],$$

also auch

$$\mathbb{1}_F E[X_m \,|\, \mathscr{F}_n] \xrightarrow{L^1} \mathbb{1}_F E[X_\infty \,|\, \mathscr{F}_n], \quad \forall\, F \in \mathscr{F},$$

denn

$$E[|\mathbb{1}_F(Y-Z)|] = E[\mathbb{1}_F|Y-Z|] \leq E[|Y-Z|].$$

Wir erhalten

$$E[\mathbb{1}_F E[X_\infty \,|\, \mathscr{F}_n]] = \lim_{m\to\infty} E[\mathbb{1}_F E[X_m \,|\, \mathscr{F}_n]] \geq E[\mathbb{1}_F X_n], \quad \forall\, F \in \mathscr{F}_n,$$

da X ein Submartingal ist und es folgt

$$E[X_\infty \,|\, \mathscr{F}_n] \geq X_n$$

Schließlich beweisen wir noch (iii) $\Rightarrow$ (i):[41] Nach Voraussetzung konvergiert $X_n \to X_\infty$ fast sicher, also insbesondere $X_n^+ \to X_\infty^+$. Weiterhin ist $(X_n^+)_{n\in\bar{\mathbb{N}}}$ gleichgradig integrierbar: In der Tat, zunächst ist auch $(X_n^+)_{n\in\bar{N}}$ ein Supermartingal, also $X_n^+ \leq E[X_\infty^+|\mathscr{F}_n]$. Damit folgt

$$\begin{aligned} E\big[(|X_n^+| - K)_+\big] = E\big[(X_n^+ - K)_+\big] &\leq E\big[(E[X_\infty^+ \,|\, \mathscr{F}_n] - K)_+\big] \\ &\leq E\big[E[(X_\infty^+ - K)_+ \,|\, \mathscr{F}_n]\big] \to 0. \end{aligned}$$

Weiterhin ist (wegen gleichgradiger Integrierbarkeit) auch $E[X_n^+] \to E[X_\infty^+]$. Da auch $E[X_n] \to E[X_\infty]$ gilt (nach Voraussetzung), erhalten wir ebenso $E[X_n^-] \to E[X_\infty^-]$.

Weiterhin folgt aus der Voraussetzung, das $X_n^- \to X_\infty^-$ f.s., was zusammen mit Konvergenz der Erwartungswerte nach Korollar 27 die L^1-Konvergenz impliziert. Damit folgt nach dem Konvergenzsatz von Vitali, Theorem 26, gleichgradige Integrierbarkeit.

□

[41] Diesen Beweis findet man typischerweise nicht in Textbüchern. Lässt man eine der Annahme fallen, gilt die Aussage nicht mehr.

Theorem 206. *Sei $\mathscr{F}_\infty = \sigma(\bigcup_{n\geq 1}\mathscr{F}_n)$, $p > 1$ und $(X_n)_{n\geq 1}$ ein Martingal so dass $\sup E[|X_n|^p] < \infty$. Dann existiert $X_\infty \in L^0(\mathscr{F}_\infty)$ mit $E[|X_\infty|^p] < \infty$, so dass*

$$X_n \xrightarrow{f.s., L^p} X_\infty.$$

Weiterhin ist $(|X_n|^p)_{n\geq 1}$ gleichgradig integrierbar.

Theorem 206 ist eine Version der berühmten **Martingalkonvergenzsätze** von Joseph L. Doob.

Beweis. Wegen $\sup E[|X_n|^p] < \infty$ ist (X_n) gleichgradig integrierbar. Nach dem vorigen Theorem gibt es X_n, s. d.

$$X_n \xrightarrow{f.s.} X_\infty.$$

Nach der Doobschen Ungleichung, Satz 201 gilt

$$E[\sup_n |X_n|^p] = \lim_{n\to\infty} E[\sup_{m\leq n} |X_m|^p] \leq \lim_n \left(\frac{p}{p-1}\right)^p E[|X_n|^p] < \infty.$$

Damit ist $\sup_n |X_n|^p]$ eine integrierbare Majorante für $(|X_n|^p)$, also ist $(|X_n|^p)$ gleichgradig integrierbar. Weiterin erhalten wir mit Fatou

$$E[|X_\infty|^p] \leq \sup_{n\to\infty} E[|X_n|^p] < \infty.$$

Nun haben wir $X_n \xrightarrow{P} X_\infty$ und $(|X_n|^p)$ ist gleichgradig integrierbar, so dass auch L_p-Konvergenz folgt. □

Supermartingale mit einer Schranke nach unten konvergieren immer, wie wir nun mit folgendem Lemma erhalten.

Lemma 207. *Ist X ein nicht-negatives Supermartingal, so existiert $X_\infty \in L^1$ mit $E[X_\infty] \leq E[X_0]$ und $X_n \xrightarrow{f.s.} X_\infty$.*

Beweis. Dann ist $-X$ ein Submartingal mit $X \leq 0$, also folgt nach dem Martingalkonvergenzsatz $X \to X_\infty$. Mit Fatou folgt

$$E[X_\infty] \leq \liminf E[X_n] \leq E[X_0]. \qquad \square$$

Beispiel 208 (Kakutanis Produkt-Martingal-Problem)**.** Seien $X_1, X_2, \dots \geq 0$ unabhängig mit $E[X_i] = 1$, $i = 1, \dots, n$ und

$$S_n = \prod_{i=1}^n X_i.$$

Dann ist

$$E[S_n \mid \mathscr{F}_{n-1}] = S_{n-1}E[X_n] = S_{n-1},$$

also (S_n) ein Martingal. Nach Lemma 207 konvergiert $S_n \to S_\infty$ f. s. Wir definieren

$$a_i = E[\sqrt{X_i}].$$

Dann ist (S_n) genau dann gleichgradig integrierbar falls $\prod_{n\geq i} a_n > 0$.

Insbesondere gilt dann $S_n \xrightarrow{L_1} S_\infty$. Ist $\prod a_i = 0$, so ist $S_n = 0$ P-f. s.

„ $\Leftarrow$ " Zunächst ist $a_i^2 = E[\sqrt{x}_i]^2 \overset{\text{Jensen}}{\leq} E[X_i] = 1$, also $a_i \leq 1$. Wir setzen $W_n = \prod_{i=1}^n \frac{\sqrt{x_i^2}}{a_i}$, so dass (W_n) ein Martingal ist

$$\Rightarrow \sup_{n\geq 1} E[W_n^2] = \sup E\Big[\prod_{i=1}^n \frac{x_i}{a_i^2}\Big] = \sup \prod_{i=1}^n E\Big[\frac{X_i}{a_i^2}\Big] \leq \frac{1}{(\prod_{n\geq 1} a_n)^2} < \infty.$$

Also ist (W_n^2) gleichgradig integrierbar nach Theorem 206. Dann gilt nach Theorem 206 $W_n \underset{f.s.}{\rightarrow} W_\infty = \frac{\prod x_i}{\prod a_i}$.

Da $\prod a_i \cdot X^2 =: f$ stetig ist, konvergiert auch

$$f(W_n) = S_n \underset{f.s.}{\rightarrow} f(W_\infty).$$

Nach Theorem 205 ist (S_n) gleichgradig integrierbar.

Theorem 209. *Sei $X \in L^1$ und $\mathscr{F}_\infty = \sigma(\bigcup_{n\geq 1} \mathscr{F}_n)$ und $\mathscr{F}_0 = \bigcap_{n\geq 1} \mathscr{F}_n$. Dann gilt*

$$E[X \mid \mathscr{F}_n] \xrightarrow[n\to\infty]{f.s.,L_1} E[X \mid \mathscr{F}_\infty], \text{ und}$$
$$E[X \mid \mathscr{F}_n] \xrightarrow[n\to 0]{f.s.,L_1} E[X \mid \mathscr{F}_0]$$

Beweis. Wir betrachten das Martingal

$$X_n = E[X \mid \mathscr{F}_n].$$

Da $E[|X_n|] \leq E[|X|] < \infty$ folgt mit dem Martingalkonvergenzsatz

$$X_n \xrightarrow{f.s.} X_\infty$$

Damit ist (X_n) gleichgradig integrierbar und somit auch

$$X_n \xrightarrow{L^1} X_\infty.$$

Der Grenzwert kann $\mathscr{F}_\infty$-messbar gewählt werden. Nun zeigen wir $X_\infty = E[X \mid \mathscr{F}_\infty]$.

Sei $F \in \mathscr{F}_m$, $m \leq n$. Dann ist

$$E[E[X \mid \mathscr{F}_n]\mathbb{1}_F] = E[X\mathbb{1}_F].$$

Mit majorisierter Konvergenz folgt $E[X_\infty \mathbb{1}_F] = E[X\mathbb{1}_F]$ also die Behauptung

Teil (ii) folgt analog. $\square$

Rückwärtsmartingale

Eine Familie von Zufallsvariable $(X_n)_{n\in-\mathbb{N}_0}$ nennen[42] wir ebenfalls einen stochastischen Prozess. Die Indizierung ist nach wie vor vorwärts, schaut man aber von der Null aus, so kehrt sich die Richtung um. Dieser Prozess hat einen natürlichen Endpunkt, nämlich X_0.

[42] $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Definition 210. Gilt $X_n \in \mathscr{F}_n$, $X_n \in L^1 \quad \forall n \in -\mathbb{N}_0$ und

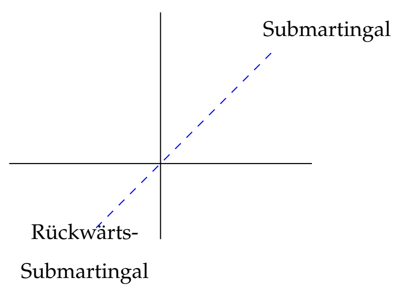
$$E[X_n \mid \mathscr{F}_{n-1}] = X_{n-1}, \quad P\text{-f. s.}, \tag{211}$$

so nennen wir $(X_n)_{n\in-\mathbb{N}_0}$ *Rückwärts-Martingal* (Rückwärts-Sub-/Super-Martingal mit $\leq$ bzw. $\geq$). .

Betrachten wir den stochastischen Prozess in umgekehrter Richtung, also $(X_{-n})_{n\in\mathbb{N}}$, so wird aus

$$\begin{aligned} E[X_{-n} \mid \mathscr{F}_{-n-1}] &= X_{-n-1}, \\ \text{also} \quad E[X_{-n} \mid \mathscr{F}_{-(n+1)}] &= X_{-(n+1)}. \end{aligned}$$

Die Interpretation Sub-/Supermartingal kehrt sich damit um.



Die zentrale Idee zeigt sich bereits im folgenden Resultat.

Lemma 212. *Sei $X = (X_n)_{n\in-\mathbb{N}_0}$ ein Rückwärtsmartingal. Dann ist X gleichgradig integrierbar.*

Beweis. Nach Definition ist für $n \leq 0$

$$X_n = E[X_0 \mid \mathscr{F}_n],$$

also $X = (E[X_0] \mid \mathscr{F}_n])_{n\in\mathbb{N}}$, und somit gleichgradig integrierbar nach Lemma 203. $\square$

Wir erhalten ebenso automatisch fast sichere Konvergenz.

Satz 213. *Sei $(X_n)_{n\in -\mathbb{N}_0}$ ein Rückwärts-Submartingal. Dann gibt es eine Zufallsvariable X mit Werten in $\mathbb{R} \cup \{-\infty\}$, so dass*

$$X_n \xrightarrow[n\to -\infty]{} X \quad P\text{-f. s.}$$

Beweis. Wir imitieren die Schritte zum Beweis des Doobschen Konvergenzssatzes. Mit $U^n_{[a,b]}$ bezeichnen wir die Anzahl der Upcrossings von $(X_m)_{n\le m\le 0}$. Dann ist $U^n_{[a,b]}$ monoton fallend (!), also

$$U_{[a,b]} := \lim_{n\to -\infty} U^n_{[a,b]}$$

eine Zufallsvariable mit Werten in $\mathbb{N} \cup \{\infty\}$. Nach der Upcrossing-Inequality gilt

$$E[U_{[a,b]}] = \lim_{n\to -\infty} E[U^n_{[a,b]}] \le \frac{E[|X_0|] + |a|}{b-a} < \infty.$$

Wie in Satz 200 folgt

$$\liminf_{n\to -\infty} X_n = \limsup_{n\to -\infty} X_n := X$$

und somit die fast sichere Konvergenz. Mit Fatou gilt: Da $(X_n^+)_{n\in\mathbb{N}}$ ebenfalls ein Submartingal ist, folgt

$$E[X^+] = E[\lim_{n\to\infty} X_n^+] \le \liminf_{n\to -\infty} E]X_n^+] \le E[X_0^+] \le E[|X_0|] < \infty.$$

Hieraus erhalten wir $P(X=\infty) = 0$. □

Theorem 214. *Sei $(X_n)_{n\in -\mathbb{N}_0}$ ein Rückwärts-Submartingal und $\mathscr{F}_{-\infty} = \bigcap_{n\in -\mathbb{N}_0} \mathscr{F}_n$ und*

$$\inf_{n\in -\mathbb{N}_0} E[X_n^-] < \infty.$$

Dann existiert eine $\mathscr{F}_{-\infty}$-messbare Zufallsvariable $X_{-\infty} \in L^1$, so dass

$$X_n \xrightarrow[n\to -\infty]{f.s., L^1} X_{-\infty}.$$

Außerdem ist $(X_n)_{\in -\mathbb{N}\cup\{-a\}}$ ein (Rückwärts-)Submartingal.

Beweis. Wir nutzten eine Doob-Meyer-Zerlegung für den Beweis. Sei dazu

$$X_n = M_n + A_n$$

mit $M_n = X_n - A_n$ und

$$A_n = \sum_{m=n+1}^{0} E[X_n - X_{n-1} \mid \mathscr{F}_{n-1}].$$

Dann ist (M_n) ein Rückwärtsmartingal:

$$\begin{aligned} E[M_n \mid \mathscr{F}_{n-1}] &= E[X_n \mid \mathscr{F}_{n-1}] - E[X_n - X_{n-1} \mid \mathscr{F}_{n-1}] - A_{n-1} \\ &= M_{n-1} \end{aligned}$$

Außerdem ist A_n monoton wachsend, so dass eine $\mathscr{F}_{-\infty}$-messbare Zufallsvariable A existiert und

$$A_n \xrightarrow{f.s.} A$$

Weiterhin ist

$$\begin{aligned} E[A] = E[\lim_{n\to-\infty} A_n] \overset{\text{monot. K.}}{=} \lim_{n\to\infty} E[A_n] &= \lim_{n\to\infty} E[\sum_{m=n+1}^{0} (X_m - X_{m-1})] \\ &= \lim_{n\to\infty} E[X_n - X_0] \\ &\le E[X_0] + \sup E[X_n^+] + \inf E[X_n^-] < \infty, \end{aligned}$$

da $E[X_n^+] \le E[X_0^+]$.

Somit ist $A \in \mathcal{L}^1$ und $A_n \xrightarrow{L^1} A$. $(E[|A - A_n|] = E[A - A_n] \to 0)$.

Ebenso ist M_n ein Rückwärtsmartingal, so dass $(M_n), (A_n)$ gleichgradig integrierbar sind und somit $(M_n + A_n)$. Da $(M_n + A_n)$ fast sicher konvergieren, konvergieren sie auch in L^1. □

Beispiel 215 (Kolmogorov's Starkes Gesetz der großen Zahlen)**.**
Wir betrachten X_n iid mit $E[|X_1|] < \infty$ und $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Ohne Beschränkung der Allgemeinheit sei $E[X_1] = 0$.

Wir setzen $\mathscr{F}_{-n} := \sigma(S_m : m \ge n)$.

Dann ist $S_{m+1} - S_m = X_{m+1}$, also

$$\mathscr{F}_{-n} = \sigma(S_n, X_{n+1}, \dots)$$

Wir erhalten

$$E[X_1 \mid \mathscr{F}_{-n}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i \mid S_n] = E[X_n \mid S_n] = S_n,$$

also $(S_n) = (E[X_1 \mid \mathscr{F}_{-n}])$ ein Rückwärtsmartingal. Nun ist $\mathscr{F}_{-\infty}$ eine terminale Sigma-Algebra, also $S_{-\infty}$ fast sicher konstant und es folgt
$S_n \to 0 \quad P$-f. s.

Das Fundamental Theorem of Asset Pricing

Ein hervorragendes Buch zu diesem Thema ist das Buch von Hans Föllmer und Alexander Schied[43] (am besten in der neuesten Edition). Wir geben nur einen kurzen Einblick – dieses und andere Resultate werden in der Vorlesung *Finanzmathematik in diskreter Zeit* behandelt (und in allgemeinerer Form in *Stochastische Integration und Finanzmathematik*).

[43] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter, Berlin, 2011

Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, P)$ und eine Filtration $\mathbb{F} = (\mathscr{F}_t)_{t=0,\dots,T}$. Eine Aktie[44] ist ein adpatierter stochastischer Prozess S. Eine Handelsstrategie ist ein vorhersehbarer Prozess $H = (H_t)_{t=1,\dots,T}$. Die Gewinne aus (selbstfinanzierendem) Handeln sind

[44] Etwas vereinfachend betrachten wir bereits diskontierte Größen.

$$G_T = \sum_{t=1}^{T} H_t \, \Delta S_t$$

mit der Aktienveränderung $\Delta S_t = S_t - S_{t-1}$.

Eine **Arbitrage** ist eine selbstfinanzierende Handelsstrategie, so dass $G_T \geq 0$ ist aber $G_T \neq 0$.

Eine Richtung des FTAP (Fundamental Theorem of Asset Pricing - Hauptsatz der Wertpapierbewertung) ist die folgende Aussage:

Theorem 216. *Existiert ein äquivalentes Martingalmaß Q, so ist der Markt frei von Arbitrage.*

Beweis. Q ist dann ein Martingalmaß, wenn S ein Q-Martingal ist. Weiterhin ist es äquivalent zu P wenn es die gleichen Nullmengen besitzt.

Angenommen H ist eine Arbitrage. Dann ist $G_T \geq 0$ und $P(G_T > 0) > 0$. Also auch $Q(G_T > 0)$.

Allerdings ist auch

$$\begin{aligned}
E_Q[G_T] &= \sum_{t=1}^{T} E_Q[H_t \, \Delta S_t] \\
&= \sum_{t=1}^{T} E_Q\big[E_Q[H_t \, \Delta S_t | \mathscr{F}_t]\big] \\
&= \sum_{t=1}^{T} E_Q\big[H_t \, E_Q[\Delta S_t | \mathscr{F}_t]\big] = 0,
\end{aligned}$$

ein Widerspruch, da $E_Q[G_T] \geq 0$ und $Q(G_T > 0)$. □

Anhang: Maßtheorie

Elementare Strukturen

Das klassische Beispiel einer σ-Algebra ist die Borel σ-Algebra $\mathscr{B}(\mathbb{R})$. Sie wird von abzählbar vielen Mengen erzeugt: etwa von den Intervallen $(a, b]$ mit rationalen Intervallgrenzen. Diese Menge ist kein Ring, sondern lediglich ein Halbring[45]

Ein *topologischer Raum*[46] $(\Omega, \mathcal{O})$ ist ein Raum Ω und einer Menge $\mathcal{O} \subset \mathscr{P}(\Omega)$, wobei die Elemente von $\mathcal{O}$ als offen bezeichnet werden, so dass

(i) $\emptyset, \Omega \in \mathcal{O}$
(ii) der endliche Schnitt von offenen Mengen ist wieder offen
(iii) die beliebige Vereinigung von offenen Mengen ist wieder offen.

[45] Ein Halbring ist durchschnittsstabil und jedes $A \backslash B$ lässt sich als endliche Summe paarweise disjunkter Mengen darstellen.
[46] $\leftarrow$ topologischer Raum $(\Omega, \mathcal{O})$

Hat man z.B. einen metrischen Raum (etwa die stetigen Funktionen über $[0, 1]$ versehen mit der Supremumsnorem), so induziert dies eine Topologie über alle offenen Kugeln.

Sei Ω eine (beliebige) Menge. Betrachten wir für einen Moment $\Omega = [0, 1]$. Wir ordnen jedem Intervall die Wahrscheinlichkeit $P((a, b]) = b - a$ zu. Angenommen [47] wir wollten dieses Maß auf alle Teilmengen von $[0, 1]$ ausdehnen, so dass (i) $P(\Omega) = 1$, P ist σ-additiv für paarweise disjunkte Mengen, so kann man zeigen, dass keine solche Erweiterung existiert! Es sind einfach zu viele Mengen. Émile Borel entdeckte, dass wir dies aber auf einer kleineren Menge tun können.

[47] Dies führt zum berühmten Banach-Tarski Paradoxon.

Mit $\mathscr{P}(\Omega)$ bezeichnen wir die Menge aller Teilmengen von Omega.

Definition A.1. Ein $\mathscr{F} \subset \mathscr{P}(\Omega)$ heißt *σ-Algebra*, falls für alle $A, A_1, A_2, \ldots \in \mathscr{F}$ gilt, dass

(i) $\emptyset \in \mathscr{F}$
(ii) $A^c \in \mathscr{F}$
(iii) $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$

Der Schnitt von (beliebig vielen) σ-Algebren ist wieder eine σ-Algebra ($\rightarrow$ Übung), so dass man für ein $C \subseteq \mathscr{P}(\Omega)$ die von $\boldsymbol{C}$

erzeugte σ-Algebra definiert durch

$$\sigma(C) = \bigcap \{\mathscr{F} \supseteq C : \mathscr{F} \text{ ist } \sigma\text{-Algebra}\}. \tag{A.2}$$

Definition A.3. Sei $(\Omega, \mathcal{O})$ ein topologischer Raum. Dann heißt $\mathscr{B}(\Omega) = \sigma(\mathcal{O})$ die *Borel-σ-Algebra* auf $(\Omega, \mathcal{O})$.

Hat $\mathcal{O}$ eine abzählbare Basis[48] C, so ist $\sigma(\mathcal{O}) = \sigma(C)$. $\to$ Übung.

Insbesondere erzeugt $C = \{(-\infty, x] : x \in \mathbb{Q}\}$ die Borel-σ-Algebra $\mathscr{B}(\mathbb{R})$. Das gilt auch für die 2-Punkt-Kompaktifizierung

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}.$$

Oft gibt es geeignete, einfache erzeugende Systeme[49]. Etwa Monotone Klassen oder Dynkin Systeme.

Definition A.4. $\mathscr{M} \subset \mathscr{P}(\Omega)$ heißt *monotone Klasse*, falls mit $A_i \in \mathscr{M}$ gilt:

(i) $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{M}$

(ii) $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathscr{M}$

Kann man in einer monotonen Klasse Komplemente bilden und ist $\Omega \in \mathscr{M}$, so ist sie ein Dynkin System.

Definition A.5. $\mathscr{D} \subset \mathscr{P}(\Omega)$ heißt *Dynkin-System*, falls für alle $A_1, A_2, A_3, \ldots \in \mathscr{D}$ gilt:

(i) $\Omega \in \mathscr{D}$

(ii) $A_1 \subseteq A_2 \Rightarrow A_2 \setminus A_1 \in \mathscr{D}$

(iii) $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$

Natürlich ist jede σ-Algebra ein Dynkin-System. Mit Dynkin-Systemen lassen sich viele Aussagen über σ-Algebren elegant beweisen: Man nutzt hierbei ($\to$ Übung), dass beliebige Schnitte von Dynkin-Systemen wieder Dynkin-Systeme sind. Hiermit können wir das von einer Menge C erzeugte Dynkin-System $\lambda(C)$ definieren[50].

Dynkin-System als wichtige Beweistechnik

(i) Angenommen $\mathscr{F} = \sigma(C)$ wobei C durchschnittsstabil ist

[48] Eine Menge C heißt Basis des topologischen Raumes $(\Omega, \mathcal{O})$, falls sich jede offene Menge als Vereinigung beliebig vieler Mengen aus C schreiben lässt.

[49] In der englischen Literatur heißt ein durchschnittsstabiles System π-System; ein Dynkin-System λ-System (siehe etwa Billingsley.

[50] Wie?

(ii) Alle Mengen in C haben die Eigenschaft die wir nachweisen wollen

(iii) Man betrachtet nun die Menge $\mathscr{D}$ aller Teilmengen von $\mathscr{F}$ mit der gewünschten Eigenschaft.

(iv) Kann man zeigen, dass $\mathscr{D}$ ein Dynkin-System ist, so ist man fertig: Nach dem nun folgenden Satz ist $C \subseteq \mathscr{D}$. Das von C erzeugte Dynkin-System $\lambda(C)$ ist gleich $\sigma(C) = \mathscr{F}$. Außerdem ist $\sigma(C) = \lambda(C) \subseteq \mathscr{D} \subseteq \mathscr{F}$ und somit $\mathscr{D} = \mathscr{F}$.

Ein Mengensystem $\mathscr{C}$ heißt *durchschnittsstabil*, fall mit $A, B \in \mathscr{C}$ auch $A \cap C \in \mathscr{C}$.

Lemma A.6. *Ist $\mathscr{D}$ Dynkin-System und $C \subseteq \mathscr{D}$ durchschnittsstabil, so gilt*

$$\sigma(C) \subseteq \mathscr{D}.$$

⇒ Jedes durchschnittsstabile Dynkin-System ist eine σ-Algebra.

Beweis. Wir definieren durch

$$\lambda(C) = \bigcap\{\mathscr{F} : C \subseteq \mathscr{F} \text{ und } \mathscr{F} \text{ Dynkin-System}\}$$

das von C erzeugte Dynkin System. Da der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist, ist $C \subseteq \lambda(C)$. Der Rest des Beweises unterteilt sich in drei Schritte:

1: $\lambda(C)$ ist durchschnittsstabil. Seien $A, B \in \lambda(C)$. Sind $A, B \in C$, so folgt auch $A \cap B \in C \subseteq \lambda(C)$. Sei nun lediglich $B \in C$. Wir setzen[51]

[51] Ein erste Anwendung der Dynkin-Beweistechnik

$$\mathscr{D}_B := \{A \subseteq \Omega : A \cap B \in \lambda(C)\}.$$

Damit ist $\mathscr{D}_B$ ein Dynkin-System, denn für $A_1, A_2, \ldots \in \mathscr{D}_B$ gilt

(i) $\Omega \in \mathscr{D}_B$

(ii) $A_1 \cap B, A_2 \cap B \in \lambda(C)$ mit $A_1 \subseteq A_2$:
$\Rightarrow A_1 \cap B \subseteq A_2 \cap B \Rightarrow (A_2 \setminus A_1) \cap B = (A_2 \cap B) \setminus (A_1 \cap B) \in \lambda(C)$.

(iii) ebenso $A_1 \cap B \subseteq A_2 \cap B \ldots$

$$\Rightarrow \Big(\bigcup_i A_i\Big) \cap B = \bigcup_i (A_i \cap B) \in \lambda(C).$$

Da $C \subset \mathscr{D}_B$ folgt nun $\lambda(C) \subseteq \mathscr{D}_B$. Wir erhalten für $A \in \lambda(C) \subseteq \mathscr{D}_B$, dass

$$A \cap B \in \lambda(C). \qquad \text{(Def. von } \mathscr{D}_B)$$

Nun setzen wir für $A \in \lambda(C)$[52]

[52] Ein zweites Mal der Dynkin-Trick.

$$\mathscr{D}_A := \{B \subseteq \Omega : A \cap B \in \lambda(C)\}.$$

Wie vorher zeigt man, dass $\mathscr{D}_A$ ein Dynkin-System ist mit $C \subset \mathscr{D}_A$ $\Rightarrow \lambda(C) \subseteq \mathscr{D}_A$, also für $A, B \in \lambda(C)$ gerade

$$A \cap B \in \lambda(C). \qquad \text{Das war 1)}$$

2: $\lambda(C)$ ist σ-Algebra. Für $A_1, A_2, \ldots \in \lambda(C)$ ist $A_i^c \in \lambda(C)$ und somit

$$\Big(\bigcup_{i=1}^n A_i\Big) = \Big(\bigcap_{i=1}^n A_i^c\Big)^c \in \lambda(C) \qquad (\lambda(c)\ \cap\text{-stabil})$$

und

$$\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty \bigcup_{i=1}^n A_i \in \lambda(C).$$

3: Abschließend erhalten wir $\sigma(C) \subseteq \sigma(\lambda(C)) = \lambda(C) \subseteq \mathscr{D}$. □

Maße

Nun kommen wir zu dem wichtigen Begriff **Maß**. [53]

[53] Die folgenden Begriffe kann man auch für Ringe und Halbringe anstelle von σ-Algebren definieren. Auch kann man signierte Maße mit möglicherweise negativen Werten betrachten.

Definition A.7. Sei $\mathscr{F}$ eine σ-Algebra und $\mu : \mathscr{F} \to \mathbb{R}_{\geq 0}$. Gilt für $(A_i) \in \mathscr{F}$ p. d.

(i) $\mu(\varnothing) = 0$

(ii) $\mu\Big(\sum_{i=1}^\infty A_i\Big) = \sum_{i=1}^\infty \mu(A_i)$

so heißt μ *Maß*. μ heißt *endlich*, falls $\mu(\Omega) < \infty$ und *σ-finit*, falls es $(A_n)_{n\in\mathbb{N}} \in \mathscr{F}$ gibt, so dass $\sum_{n\in\mathbb{N}} A_n = \Omega$ und $\mu(A_n) < \infty$, für alle $n \in \mathbb{N}$.

Die Eigenschaft (ii) heißt σ-Additivität. Interessant ist es, sich Mengenfunktionen anzuschauen, wo in (ii) $\leq$ auftaucht, μ also nur sub-additiv (monoton) ist. Dann heißt μ **Kapazität** oder – was wir später noch kennen lernen – äußeres Maß.

Endliche, additive Mengenfunktionen heißen oft **Inhalt**, σ-additive Mengenfunktionen, die nicht auf σ-Algebren definiert sind **Prämaß**.

Die Fortsetzung von Maßen ist ein wichtiges Hilfsmittel: Das Lebesgue-Maß etwa definiert man durch

$$\mu([a, b]) := b - a \qquad a, b \in \mathbb{Q}$$

und setzt dann geeignet auf $\mathscr{D}(\mathbb{Q})$ fort. Wie das technisch funktioniert, soll nun erläutert werden.

Definition A.8. Ein nicht-leeres Mengensystem $\mathscr{H} \subseteq \mathscr{P}(\Omega)$ heißt *Halbring*, falls

(i) $\varnothing \in \mathscr{H}$
(ii) $A, B \in \mathscr{H} \Rightarrow A \cap B \in \mathscr{H}$ (durchschnittsstabil)
(iii) Für $A, B \in \mathscr{H}$ gibt es p. d. Mengen $C_1, \dots, C_n \in \mathscr{H}$, s.d.

$$B \setminus A = \sum_{i=1}^{n} C_i.$$

Ein nicht-leeres Mengensystem $\mathscr{R} \subseteq \mathscr{P}(\Omega)$ heißt *Ring*, falls

(i) $\varnothing \in \mathscr{R}$
(ii) $A, B \in \mathscr{R} \Rightarrow A \cup B \in \mathscr{R}$
(iii) $A, B \in \mathscr{R} \Rightarrow B \setminus A \in \mathscr{R}$.

Ein typisches Beispiel für einen Halbring ist

$$\{(a, b] : a \leq b \in \mathbb{Q}\}.$$

Dieser erzeugt $\mathscr{B}(\mathbb{R})$.

Ein Ring heißt Algebra, falls $\Omega \in \mathscr{R}$. Die σ-Algebra erfüllt zusätzlich die σ-Additivität.

Wir beginnen mit einem kleinen Hilfsresultat. Wie auch im Folgenden nutzen wir $\sum$ um eine Summe über paarweise disjunkte Mengen darzustellen.

Lemma A.9. *Seien $A_1, \dots, A_n \in \mathscr{H}$ und $B_n = A_n \setminus \left(\cup_{i=1}^{n-1} A_i\right)$. Dann ist*

$$\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} B_i.$$

Außerdem gibt es $D_1, \dots, D_k \in \mathscr{H}$ so dass

$$B_n = \sum_{i=1}^{k} D_i.$$

Beweis. Für $n = 2$ folgt dies aus der Halbring-Eigenschaft (ii). Gilt die Behauptung für ein n, so ist

$$\begin{aligned} B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^{n} A_i &= \Bigg(\underbrace{A_{n+1} \setminus \bigcup_{i=2}^{n} A_i}_{=\sum_{i=1}^{k} D_i \text{ nach Vorr.}} \Bigg) \setminus A_1 \\ &= \sum_{i=1}^{k} D_i \setminus A_1 \end{aligned}$$

Da $\mathscr{H}$ ein Halbring ist, gibt es $(E_{ij})_{i=1,\dots,k}$ so dass $D_i \setminus A_1 =$

$\sum_{j=1}^{k_i} E_{ij}$, also ist

$$B_{n+1} = \sum_{i=1}^{k^{(n+1)}} \sum_{j=1}^{k_i} E_{ij}.$$

□

Als wichtige Bemerkung erhalten wir auch

$$A_{n+1} = B_{n+1} + A_{n+1} \cap \Big(\underbrace{\underbrace{\bigcup_{m=1}^{n} A_m}_{=\sum_{m=1}^{n} B_m}}_{=\sum_{m=1}^{n} \sum_{i=1}^{k(m)} \sum_{j=1}^{k_j} E_{ij}} \Big)$$

Lemma A.10. *Sei $\mathscr{H}$ ein Halbring mit $\mu : \mathscr{H} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ additiv. Dann gilt:*

(i) μ ist monoton und subadditiv
(ii) μ ist σ-additiv $\Leftrightarrow$ μ ist σ-subadditiv

Beweis. Wir zeigen zunächst, dass μ monoton ist. Seien $A, B \in \mathscr{H}$, $A \subseteq B \Rightarrow$

$$B = A + B \setminus A = \underbrace{A}_{\in \mathscr{H}} + \sum_{i=1}^{k} \underbrace{D_i}_{\in \mathscr{H}},$$

also ist $\mu(B) = \mu(A) + \sum_{i=1}^{k} \mu(D_i) \geq \mu(A)$. [54]

[54] Denn $\mu(D_i) \geq 0$.

Nun zeigen wir die Subadditivität: Seien $A_1, \dots, A_n, \bigcup_{i=1}^{n} A_i \in \mathscr{H}$. Mit Lemma A.9 ist

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \mu\Big(\sum_{i=1}^{n} B_i\Big) = \mu\Big(\sum_{i=1}^{n}\sum_{j=1}^{k_i} D_{ij}\Big) = \sum_{i=1}^{n}\sum_{j=1}^{k_i} \mu(D_{ij})$$
$$\leq \sum_{i=1}^{n}\sum_{j=1}^{k_i} \mu(D_{ij}) + \sum \cdots = \sum_{i=1}^{n} \mu(A_i) \quad \text{(A.11)}$$

Nun folgt μ σ-additiv $\Rightarrow$ μ σ-subadditiv direkt wie (A.11).
Die Rückrichtung zeigen wir etwas später (ÜA).

□

Beispiel A.12. Die Rückrichtung gilt **nicht** wenn $\mathscr{H}$ kein Halbring ist: Man wähle zum Beispiel[55] $\mathscr{H} = \{n : n \in \mathbb{N}\} \cup \mathbb{N}$. Für zwei unterschiedliche Mengen A und B ist dann $\Rightarrow$ $A \cap B$ entweder leer oder enthält nur ein Element. Wir wählen ein σ-subadditives Maß durch

[55] Wieso ist das kein Halbring?

$$\mu(A) := \begin{cases} = 1 & A = \mathbb{N} \\ = n^{-2} \quad A & = \{n\} \end{cases}.$$

Allerdings ist dann

$$P\Big(\sum_{n\in\mathbb{N}}\{n\}\Big) = \sum \frac{1}{n^2} \neq 1 = P(\mathbb{N})$$

und μ ist demnach nicht σ-additiv.

Ist $\mathscr{H}$ ein Halbring, so nennen wir

$$\mathscr{R}(\mathscr{H}) = \Big\{\sum_{i=1}^{n} A_i : A_1,\ldots,A_n \in \mathscr{H} \text{ p. d.}, n \in \mathbb{N}\Big\}$$

den von $\mathscr{H}$ **erzeugten Ring**. Man kann diesen Ring auch über die folgende Eigenschaft definieren: Beliebige Schnitte von Halbringen sind wieder Halbringe (ÜA). Dann kann man $\mathscr{R}(\mathscr{H})$ auch wie in (2) definieren. Zeigen Sie, dass die beiden Definitionen übereinstimmen.

Lemma A.13. *Sei $\mathscr{H}$ ein Halbring mit $\mu : \mathscr{H} \to \mathbb{R}_{\geq 0}$ endlich und additiv. Auf $\mathscr{R}(\mathscr{H})$ definieren wir*

$$\widetilde{\mu}\Big(\sum_{i=1}^{n} A_i\Big) := \sum_{i=1}^{n} \mu(A_i)$$

für p. d. $A_1,\ldots,A_n \in \mathscr{H}$. Dann ist $\widetilde{\mu}$ die einzige *additive Fortsetzung auf $\mathscr{R}$ die auf $\mathscr{H}$ mit μ übereinstimmt. Außerdem gilt: $\widetilde{\mu}$ σ-additiv $\Leftrightarrow$ μ σ-additiv.*

Beweis. Es ist nur zu zeigen, dass $\widetilde{\mu}$ wohldefiniert ist. Seien dazu $A_1,\ldots,A_m, B_1,\ldots,B_n$ zwei paarweise disjunkte Darstellung der gleichen Menge, also $\sum_{i=1}^{n} A_i = \sum_{j=1}^{n} B_i$. Dann folgt

$$A_\mathrm{i} = \sum_{\mathrm{j}=1}^{n} A_\mathrm{i} \cap B_\mathrm{j}, \quad B_\mathrm{j} = \sum_{\mathrm{i}=1}^{m} B_\mathrm{j} \cap A_\mathrm{i},$$

$$\sum_{i=1}^{m} \mu(A_i) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mu(A_i \cap B_j) = \sum_{j=1}^{n} \mu(B_j)$$

□

Für eine monotone Klasse $\mathscr{M} \subseteq \mathscr{P}(\Omega)$ sagen wir $\mu : \mathscr{M} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ ist

- **stetig von unten**, falls $\mu\big(\bigcup_{i=1}^{\infty} A_i\big) = \lim_{i\to\infty} \mu(A_i)$ für $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{M}$ und $\bigcup_{i=1}^{\infty} A_i \in \mathscr{M}$
- mit $\mu(A_1) < \infty$ **stetig von oben**, falls $\mu(\bigcap_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \mu(A_i)$ für $A_1 \supseteq A_2 \supseteq$
- mit $\mu(A_1) < \infty$ **stetig von oben in** $\varnothing$, falls die obige Eigenschaft $\forall \bigcap_{i=1}^{\infty} A_i = \varnothing$ gilt.

Satz A.14. *Sei $\mathscr{R}$ ein Ring und $\mu : \mathscr{R} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ additiv und $\mu(\emptyset) = 0$,*

(i) μ *ist* σ*-additiv*
(ii) μ *ist* σ*-subadditiv*
(iii) μ *ist stetig von unten*
(iv) μ *ist stetig von oben in* $\emptyset$
(v) μ *ist stetig von oben*

Dann gilt: (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv) $\Leftrightarrow$ (v) und (iv) $\Rightarrow$ (iii) falls $\mu(A) < \infty \ \forall A \in \mathscr{R}$.

Ist μ ein endliches Maß, so sind alle Bedingungen äquivalent: Aus Stetigkeit von oben in $\emptyset$ folgt bereits σ-Additivität.

Beweis.

(i) $\Rightarrow$ (ii): Lemma A.10

(ii) $\Rightarrow$ (i): Seien $A_1, A_2, \ldots \in \mathscr{R}$, p. d. so dass $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$

$$\sum_{i=1}^\infty \mu(A_i) = \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i) = \lim_{n\to\infty} \mu\Big(\sum_{i=1}^n A_i\Big) \overset{\mu \text{ monoton}}{\leq} \mu\Big(\sum_{i=1}^\infty A_i\Big) \leq \sum_{i=1}^\infty \mu(A_i)$$

hierbei ist $\sum_{i=1}^n A_i \in \mathscr{R}$, da $\mathscr{R}$ Ring

(i) $\Rightarrow$ (iii): Sei also $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{R}$ und $\sum_{i=1}^\infty A_i \in \mathscr{R}$. Nach Lemma A.9 ist $\cup_{i=1}^\infty A_i = \sum_{i=1}^\infty B_i$ mit p.d. (B_i). Wegen der σ-Additivität (i) ist also

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu\big(\sum_{i=1}^n B_i\big) = \mu\big(\sum_{i=1}^\infty B_i\big) = \mu\big(\bigcup_{i=1}^\infty A_i\big).$$

(iii) $\Rightarrow$ (i): Seien $A_1, A_2, \ldots \in \mathscr{R}$ p. d. und $\bigcup_{i=1}^\infty A_i \in \mathscr{R}$. Dann ist

$$\sum_{i=1}^n A_i \in \mathscr{R}, \text{ monoton wachsend und } \mu\Big(\sum_{i=1}^\infty A_i\Big) = \lim_{n\to\infty} \mu\Big(\sum_{i=1}^n A_i\Big) = \sum_{i=1}^\infty \mu(A_i)$$

(iii) $\Rightarrow$ (iv): Seien $A_1 \supseteq A_2 \supseteq \ldots \in \mathscr{R}$ mit $\bigcap_{i=1}^\infty A_i = \emptyset$ und $B_n = A_1 \setminus A_n \Rightarrow \emptyset = B_1 \subseteq B_2 \subseteq \ldots$, $\bigcup_{i=1}^\infty B_i = A_1$ (da $\bigcap A_i = \emptyset$) und

$$\begin{aligned}\mu(A_1) = \lim_{n\to\infty} \mu(B_n) &= \lim_{n\to\infty} \mu(A_1) - \mu(A_n) \\ &= \mu(A_1) - \lim_{n-\infty} \mu(A_n),\end{aligned}$$

so dass $\lim_{n\to\infty} \mu(A_n) = 0$ folgt.

(iv) $\Rightarrow$ (v): Seien $A_1 \supseteq A_2 \supseteq \cdots \supseteq\in \mathscr{R}$ und $A = \bigcap_{n=1}^\infty A_n \in \mathscr{R}$. Setze $B_n = A_n \setminus A$ und nutze (iv).

(v) $\Rightarrow$ (iv): klar

(iv) $\Rightarrow$ (iii): (mit μ endlich) Seien also $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{R}$ und $\bigcup_{i=1}^{\infty} A_i \in \mathscr{R}$. Setze $B_n = \bigcup_{i=1}^{\infty} A_i \setminus A_n$, so dass $\bigcap_{n=1}^{\infty} B_n = \varnothing$, also

$$0 = \lim_{n\to\infty} \mu(B_n) = \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) - \lim_{n\to\infty} \mu(A_i) \tag{A.15}$$

und somit (da beide Summanden endlich sind) $\mu(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(A_i)$.

□

Die Fortsetzung von Maßen

In diesem Abschnitt wollen wir nun von einem Halbring auf eine σ-Algebra einen Inhalt geeignet zu einem Maß fortsetzen.

Satz A.16 (Eindeutigkeit). *Sei $\mathscr{C} \subset \mathscr{P}(\Omega)$ durchschnittsstabil, es gebe $C_1 \subseteq C_2 \subseteq \cdots \in \mathscr{C}$, so dass $\bigcup_{n=1}^{\infty} C_n = \Omega$ und $\mathscr{F} = \sigma(\mathscr{C})$. Weiterhin seien $\mu, \nu : \mathscr{F} \to \mathbb{R}_{\geq 0}$ Maße, so dass $\mu(C_i) < \infty$.*

Dann ist

$$\mu = \nu \Leftrightarrow \mu(A) = \nu(A) \quad \forall A \in \mathscr{C}.$$

Beweis:

„$\Rightarrow$“ klar

„$\Leftarrow$“ Sei $C \in \mathscr{C}$ mit $\mu(C) = \nu(C) < \infty$ und

$$\mathscr{D}_C := \{A \in \mathscr{F} : \mu(A \cap C) = \nu(A \cap C\}\} \subseteq \mathscr{C}.$$

Damit ist $\mathscr{D}_C$ Dynkin-System:

(i) $\Omega \in \mathscr{D}_C$,
(ii) $A, B \in \mathscr{D}_C$,

$$\begin{aligned} A \subseteq B \Rightarrow \mu((B \setminus A) \cap C) &= \mu(B \cap C) - \mu(A \cap C) \\ &= \nu(B \cap C) - \nu(A \cap C) \\ &= \nu((B \setminus A) \cap C), \quad \text{also } B \setminus A \in \mathscr{D}_C. \end{aligned}$$

(iii) Ist $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{D}_C$ mit $\bigcup_{i=1}^{\infty} A_i \in \mathscr{D}_C$, so folgt wegen der σ-Stetigkeit von μ, ν

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i \cap C\Big) = \lim_{n\to\infty} \mu(A_n \cap C) = \lim_{n\to\infty} \nu(A_n \cap C) = \nu\Big(\bigcup_{i=1}^{\infty} A_i\Big)$$

Dann ist $\mathscr{F} = \sigma(\mathscr{C}) = \mathscr{D}_C$.

Nun betrachten wir $A \in \mathscr{F} \Rightarrow A \in \mathscr{D}_{C_n}$ und somit

$$\mu(A \cap C_n) = \nu(A \cap C_n), \text{ also}$$

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap C_n) = \lim_{n\to\infty} \nu(A \cap C_n) = \nu(A). \qquad \square$$

Wir kommen zu den von C. Carathéodory (1873–1950) eingeführten, zentralen Begriff von einem äußeren Maß. Wie bereits erwähnt, wird lediglich die σ-Additivität durch die σ-Subadditivtät ersetzt. In der folgenden Definition wird aber auch die Wortgebung äußeres Maß klar, vergleiche Satz A.19. Alternativ wird für subadditive Mengenfunktionen (auf σ-Algebren) auch der Begriff Kapazität verwendet.

Definition A.17. Ein $\eta : \mathscr{P}(\Omega) \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ heißt *äußeres Maß*, falls

(i) $\eta(\emptyset) = 0$

(ii) $A \subset B \subset \Omega \Rightarrow \eta(A) \leq \eta(B)$ (Monotonie)

(iii) Für $A_1, A_2, \dots \in \mathscr{P}(\Omega)$ gilt
$$\eta\Big(\bigcup_{i=1}^{\infty} A_1\Big) \leq \sum_{i=1}^{\infty} \eta(A_i) \qquad (\sigma\text{-Subadditivität})$$

Ein äußeres Maß ist natürlich auch subadditiv. Da es auf allen Teilmengen von Ω definiert ist, gibt es zunächst keinen vernünftigen Begriff der Meßbarkeit.

Definition A.18. Ist η ein äußeres Maß und $A \subseteq \Omega$, so heißt A *σ-messbar*, falls für alle $B \subseteq \Omega$

$$\eta(B) \geq \eta(B \cap A) + \eta(B \cap A^c).$$

Damit ist eine Menge A genau dann messbar, wenn sie **jede** Menge $B \subseteq \Omega$ zerlegt in die disjunkten Mengen $B \cap A$, $B \cap A^c$, auf denen sich η **additiv** verhält.

Satz A.19. *Sei $\mathscr{H}$ ein Halbring und μ ein* Inhalt *auf $\mathscr{H}$. Setze für $A \subseteq \Omega$ ($\inf \emptyset = \infty$)*

$$\mu^*(A) := \inf\Big\{\sum_{n=1}^{\infty} \mu(A_n) : A_n \in \mathscr{H}, A \subseteq \bigcup_{n=1}^{\infty} A_n\Big\}. \tag{A.20}$$

Dann ist μ^* äußeres Maß.

$$\text{(A.20)} \quad \Leftrightarrow \qquad \mu^*(A) = \inf\Big\{ \sum_{i=1}^{\infty} \mu(A_n) : A_n \in \mathscr{H} \text{ p. d.}, A \subseteq \sum_{n=1}^{\infty} A_n \Big\}$$

Beweis. Offensichtlich ist μ^* monoton und $\mu^*(\varnothing) = \mu(\varnothing) = 0$.

Seien $A_1, A_2, \cdots \subseteq \Omega$. Ist $\mu(A_n) = \infty$ für ein $n \in \mathbb{N}$, so sind wir fertig.

Sei also $\mu^*(A_n) < \infty \forall n$ und $\varepsilon > 0$. Dann gibt es (Infimum!) für jedes $n \in \mathbb{N}$ eine Folge $(B_{n,k})_{k \leq n} \subset \mathscr{H}$, so dass $A_n \subseteq \bigcup_{k=1}^{\infty} B_{n,k}$ und

$$\sum_{k=1}^{\infty} \mu(B_{n,k}) \leq \mu^*(A_n) + \varepsilon \cdot 2^{-n}.$$

Nun ist $(B_{n,k})$ eine abzählbare Familie von Mengen aus $\mathscr{H}$ und

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} B_{n,k}, \text{ also}$$

$$\begin{aligned}\mu^*\Big(\bigcup_{n=1}^{\infty}\Big) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu(B_{n,k}) &\leq \sum_{n=1}^{\infty} (\mu^*(A_n) + e \cdot 2^{-n}) \\ &= \sum_{n=1}^{\infty} \mu^*(A_n) + 2\varepsilon.\end{aligned}$$

Da ε beliebig war, ist μ^* äußeres Maß. □

Satz A.21. *Sei $\mu^* : \mathscr{P}(\Omega) \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ ein äußeres Maß. Dann ist*

$$\mathscr{F}^* := \{A \in \mathscr{P}(\Omega) : A \text{ ist } \mu^*\text{-messbar}\}$$

eine σ-Algebra und $\mu^*\big|_{\mathscr{F}^*}$ ein Maß.

Beweis. (i): Wir zeigen zunächst, dass $\mathscr{F}^*$ eine Algebra ist.

$\Omega \in \mathscr{F}^*$ und $A \in \mathscr{F}^* \Rightarrow A^c \in \mathscr{F}^*$ (direkt aus Def. von μ^*-messbar)

Seien $A, B \in \mathscr{F}^*$ und $C \subset \Omega$. Dann ist

$$\begin{aligned}\mu^*(C) &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c) \qquad (A \in \mathscr{F}*) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c \cap B) + \mu^*(C \cap A^c \cap B^c) \qquad (B \in \mathscr{F}^*) \\ &\geq \mu^*(\underbrace{(C \cap A)}_{=(C \cup A \cup C \cap A \cap B)} \cup (C \cap A^c \cap B)) + \mu^*(C \cap (A \cup B)^c) \\ &= \mu^*(C \cap A \cup C \cap B) + \mu^*(C \cap (A \cup B)^c)\end{aligned}$$

und somit sind $A \cup B \in \mathscr{F}^*$.

(ii) Nun zeigen wir: Sind $(A_n)_{n\geq 1} \in \mathscr{F}^*$ p. d. $\Rightarrow A = \bigcup_{i=1}^\infty \in \mathscr{F}^*$ und μ^* ist σ-additiv

Zunächst folgt $\sum_{i=1}^n A_i \in \mathscr{F}^*$, also wegen $\mu^*(A_i + A_j) = \mu^*(A_i) + \mu^*(A_j)$, denn für jedes $c \leq \Omega$:

$$\begin{aligned}\mu^*(C) &\geq \mu^*\Big(C \cap \Big(\sum_{i=1}^n A_i\Big)\Big) + \mu^*\Big(C \cap \Big(\sum_{i=1}^n A_i\Big)^c\Big) \\ &\geq \sum_{i=1}^n \mu^*(C \cap A_i) + \mu^*(C \cap A^c). \qquad \text{(Monotonie \& Additivität)}\end{aligned}$$

Dies gilt $\forall n$, also folgt

$$\begin{aligned}\mu^*(C) &\geq \sum_{i=1}^\infty \mu^*(C \cap A_i) + \mu^*(C \cap A^c) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c) \geq \mu^*(C)\end{aligned}$$

wegen der Sub-σ-Additivität. Es folgt $A \in \mathscr{F}^*$ und mit $A = C$ die σ-Additivität von μ^*. □

Theorem A.22 (Fortsetzungssatz)**.** *Sei $\mathscr{H}$ ein Halbring und $\mu : \mathscr{H} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ ein Inhalt.*

(i) Dann sind alle Mengen aus $\mathscr{H}$ μ^*-messbar.
(ii) Ist μ σ-additiv, so gilt $\mu^*\big|_{\mathscr{H}} = \mu$.
(iii) Ist μ nicht σ-additiv, so gibt es $A \in \mathscr{H}$, so dass $\mu^*(A) < \mu(A)$.

Insbesondere ist $\mu^*\big|_{\mathscr{F}^*}$ Maß und damit auch $\mu^*\big|_{\sigma(\mathscr{H})}$.

Beweis. (i) Sei $A \in \mathscr{H}$ und $C \subseteq \Omega$ mit $\mu^*(C) < \infty$ (für $\mu^*(C) = \infty$ ist nicht zu zeigen) und $(B_n)_{n\geq 1} \subseteq \mathscr{R}(\mathscr{H})$ so, dass $C \subseteq \bigcup_{i=1}^\infty B_n$ und $\mu^*(C) + \varepsilon \geq \sum \widetilde{\mu}(B_n)$ (existiert, da $\mu^*(c) < \infty$).

Wir betrachten den auf $\mathscr{R}(\mathscr{H})$ erzeugten Inhalt $\widetilde{\mu}$.

$$\begin{aligned}\mu^*(c) + \varepsilon \geq \sum_{n=1}^\infty \widetilde{\mu}(B_n) &= \sum_{n=1}^\infty \widetilde{\mu}(B_n \cap A) + \sum_{n=1}^\infty \widetilde{\mu}(B_n \cap A^c) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c).\end{aligned}$$

Da $\varepsilon > 0$ beliebig, folgt $A \in \mathscr{F}^*$.

(ii) Sei $A \in \mathscr{H}$. Für jedes $\varepsilon > 0$ gibt es $(A_n)_{n\geq 1} \subset \mathscr{H}$, so dass $A^c = \sum_{n\geq 1} A_n$ und (Def. μ^*)

$$\mu^*(A) + \varepsilon \geq \sum_{n\geq 1} \mu(A_n) \underset{\sigma\text{-Sub.-Add.}}{\geq} \mu\Big(\bigcup_{n\geq 1} A_n\Big) \geq \mu(A) \geq \mu^*(A)$$

und es folgt $\mu^*(A) = \mu(A)$.

(iii) Ist μ nicht σ-additiv, so gibt es p. d. $(A_n) \subseteq \mathscr{H}$, $A = A_n \in \mathscr{H}$ mit

$$\mu\Big(\sum_{i=1}^{\infty} A_n\Big) \neq \sum_{i=1}^{\infty} \mu(A_n).$$

Allerdings ist stets (μ Inhalt) $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\Big(\sum_{n=1}^{\infty} A_n\Big)$, also $\sum_{n=1}^{\infty} \mu(A_n) < \mu(A)$. Nun ist $\mu^*(A_n) = \mu(A)$ und μ^* σ-additiv, also $\sum_{n=1}^{\infty} \mu^{\infty}(A_n) = \mu^*(A)$. □

Man kann noch zeigen, dass

$$\mathscr{F}^* = \{A \setminus N : A \in \sigma(\mathscr{H}), N \in \Omega, \mu^*(N) = 0\}.$$

Maße auf $\mathbb{R}$

Wir können nun das Lebesgue-Maß λ eindeutig auf $\mathscr{B}(\mathbb{R})$ definieren durch

$$\lambda((a,b]) = b - a, \qquad a, b \in \mathbb{Q} \text{ mit } a \leq b.$$

Ebenso erhalten wir eine Charakterisierung aller Wahrscheinlichkeitsmaße auf $\mathscr{B}(\mathbb{R})$.

Satz A.23. *Eine Funktion $P : \mathscr{B}(\mathbb{R}) \to [0,1]$ ist ein Wahrscheinlichkeitsmaß genau dann, wenn es eine wachsende rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ gibt, so dass*

$$P((a,b]) = F(b) - F(a), \qquad a, b \in \mathbb{Q}, a \leq b.$$

Offensichtlich ist P eindeutig durch F bestimmt! Den Beweis führen wir als Übungsaufgabe.

Für uns wichtig werden Verteilungen, also Bildmaße sein.

Definition A.24. Ein Tripel $(\Omega, \mathscr{F}, \mu)$ heißt *Maßraum*, falls $\mathscr{F}$ σ-Algebra und μ Maß auf $(\Omega, \mathscr{F})$ ist.

Definition A.25. Sei $(\Omega, \mathscr{F}, \mu)$ ein Maßraum, $\mathscr{F}'$ eine σ-Algebra auf Ω'. Eine Funktion $f : \Omega \to \Omega'$ heißt *$\mathscr{F}$-$\mathscr{F}'$-messbar*, falls

$$f^{-1}(A) \in \mathscr{F} \text{ für alle } A \in \mathscr{F}'.$$

Für ein solches f heißt $f_*\mu : \mathscr{F}' \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, definiert durch

$$f_*\mu(A') = \mu(f^{-1}(A')) = \mu(\omega \in \Omega : f(\omega) \in A'), \qquad A' \in \mathscr{F}'$$

Bildmaß von μ unter f.

Im Englischen nennt man das Bildmaß *push-forward*. Sind die σ-Algebren aus dem Kontext klar, so nennt man f einfach meßbar.

Man zeigt leicht, dass $f_*\mu$ wieder ein Maß ist. Es ist die so gennte *Verteilung* der Zufallsvariablen f. Die **von f erzeugte** σ-Algebra

$$\sigma(f) := f^{-1}(\mathscr{F}')$$

ist in der Tat eine σ-Algebra. Ebenso gilt für $\sigma(\mathscr{C}') = \mathscr{F}'$, dass

$$\sigma(f^{-1}(\mathscr{C}')) = f^{-1}(\sigma(\mathscr{C}')).$$

Ist $(\Omega', \mathscr{F}') = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, so heißt f **reellwertig** und ist f $\mathscr{F}$-$\mathscr{B}(\mathbb{R})$-messbar, so nennen wir f **Borel-messbar**.

Messbarkeit ist ein wichtiger Begriff, so dass wir ein paar Rechenregeln wiederholen. Wir schreiben $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Für den Begriff *Stetigkeit* reicht es, auf topologischen Räumen zu arbeiten: Dort sind Funktionen stetig, wenn die Urbilder offener Mengen wieder offen sind.

Lemma A.26. *Seien $(\Omega, \mathscr{F})$, $(\Omega', \mathscr{F}')$, $(\Omega'', \mathscr{F}'')$ Maßräume und $f : \Omega \to \Omega'$, sowie $g : \Omega' \to \Omega''$.*

(i) $\mathscr{F}' = \sigma(\mathscr{C}') \Rightarrow f$ *messbar* $\Leftrightarrow f^{-1}(\mathscr{C}') \subseteq \mathscr{F}$

(ii) f, g *messbar* $\Rightarrow f \circ g$ *messbar*

(iii) *stetige Abbildungen sind messbar bzgl. der Borel-σ-Algebra.*

(iv) *Eine reellwertige Funktion f ist genau dann messbar, falls*

$$\{\omega : f(\omega) \le x\} \in \mathscr{F} \qquad \forall x \in \mathbb{Q}$$

(v) $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ *ist messbar falls* $A_1, \dots, A_n \in \mathscr{F}$.

(vi) $f, g : \Omega \to \overline{\mathbb{R}}$ *messbar* $\Rightarrow f \cdot g,\ a \cdot f + b \cdot g,\ \frac{f}{g}\mathbb{1}_{\{g \neq 0\}}$ *messbar* $(a, b \in \mathbb{R})$

(vii) *Sind* $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ *messbar, so auch*

$$\sup_n f_n, \inf_n f_n, \lim_n \sup f_n, \lim_n \inf f_n.$$

Am interessantesten ist wohl der Beweis für die Messbarkeit von $\sup_n f_n$:

$$\{\omega : \sup f_n(\omega) \le x\} = \bigcap_{n \ge 1} \underbrace{\{\omega : f_n(\omega) \le x\}}_{\in \mathscr{F}} \in \mathscr{F}.$$

Eine zentrale Eigenschafte: Meßbare Funktionen sind monoton durch einfache Funktionen approximierbar.

Eine wichtige Eigenschaft messbarer Funktionen ist, dass sie stets durch einfache Funktionen approximierbar sind: Wir betrachten $f \ge 0$ und setzen

$$A_{j,n} = \begin{cases} \{\frac{j}{2^n} < f \le \frac{j+1}{2^n}\}, & j = 0, \dots, n \cdot 2^n - 1 \\ \{f \ge n\}, & j = n \cdot 2^n. \end{cases}$$

Unsere Apprxoimation ist

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} \mathbb{1}_{\{A_{j,n}\}}.$$

Dann gilt $f_n \uparrow f$. Dies ist der Schlüssel zum Integral! Wir definieren für einfache Funktionen

$$\int f d\mu = \int \sum_{i=1}^{n} c_i \mathbb{1}_{A_i} d\mu = \sum_{i=1}^{n} c_i \int_{A_i} d\mu = \sum_{i=1}^{n} c_i \mu(A_i)$$

und für messbare, nicht-negative Funktionen und $f_n \uparrow f$

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu = \sup\Big\{ \int g d\mu : g \text{ einfach und } 0 \leq g \leq f \Big\}.$$

Dieses Integral können wir fortsetzten, falls mindestens ein Integral $\int f^+ d\mu, \int f^- d\mu$ endlich ist. Dann definieren wir

$$\mathcal{L}^1(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{Q}} : \int |f| d\mu < \infty \Big\} \qquad \text{(und analog } \mathcal{L}^p)$$

Leider hat dieser Raum schlechte Trennungseigenschaften: Gibt es eine nichtleere Null-Menge, so gibt es immer verschiedene meßbare Funktionen, die den Abstand Null haben. Um das zu beheben betrachtet man Äquivalenzklassen, siehe auch Abschnitt VI.2 in [56].

Auf den messbaren Funktionen kann man folgende Äquivalenzrelation einführen: $f \sim g$ falls $P(f = g) = 1$. Die hierdurch erhaltenen Äquivalenzklassen definieren die L^p-Räume. In der Notation machen wir dies durch die Unterscheidung $\mathcal{L}$ und L kenntlich. Im Folgenden kürzt f. s. fast sicher ab, etwa $f \leq g$ f. s. ist gleichbedeutent mit $\mu(\omega \in \Omega : f > g) = 0$. Meist ist dies auch aus dem Kontext klar und wir lassen das f.s. weg.

Natürlich ist

$$\int c \mathbb{1}_A d\mu = c\mu(A).$$

und hieraus entsteht sofort das Integral für einfache Funktionen.

[56] Jürgen Elstrodt. *Maß-und Integrationstheorie*. Springer, 2018

Die L^p-Räume sind die Äquivalenzklassen der integrierbaren Funktionen. Für $p \geq 1$ erhält man einen metrischen Raum der vollständig ist (Banachraum), mit $p = 2$ sogar einen Hilbertraum.

Satz A.27. *Erste Eigenschaften:*

- *(i)* $f \leq g$ *(f. s.)* $\Rightarrow \int f d\mu \leq \int g d\mu$
- *(ii)* $|\int f d\mu| \leq \int |f| d\mu$ *(Dreiecksungleichung)*
- *(iii)* $\int (af + bg) d\mu = \int f d\mu + b \int g d\mu$ *(Linearität)*
- *(iv)* $f = 0$ *(f. s.)* $\Rightarrow \int f d\mu = 0$
- *(v)* $\int f d\mu < \infty \Rightarrow f < \infty$ *(f.s.)*
- *(vi)* $f_n \uparrow f \Rightarrow \int f_n d\mu \uparrow \int f d\mu.$ *(Monotone Konvergenz)*

Als Übung beweisen wir den Substitutionssatz

Satz A.28. *$g \in \mathcal{L}^1(f_*\mu) \Rightarrow g \circ f \in \mathcal{L}^1(\mu)$ und*

$$\int g \circ f \, d\mu = \int g \, d(f_*\mu).$$

Beweis. Es genügt, die Aussage für einfache nicht-negative g zu zeigen. Der allgemeine Fall folgt durch Approximation.

Sei $g = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \Rightarrow g \circ f = \sum_{i=1}^n c_i \mathbb{1}_{\{f \in A_i\}}$, also

$$\int g \circ f d\mu = \sum_{i=1}^n c_i \, \mu(f \in A_i) = \sum_{i=12}^n c_i \, f_*\mu(A_i) = \int g \, d(f_*\mu). \qquad \square$$

So wird man viele Aussage für Integrale beweisen können: Zunächst die Aussagen auf einfachen, nicht negativen Funktionen zeigen und dann durch Approximation das Resultat auf die Integrale übertragen.

Äußerst wichtig sind die folgenden Konvergenzsätze.

Theorem A.29 (Monotone Konvergenz). *Sei $(\Omega, \mathscr{F}, \mu)$ Maßraum, $f_n \geq 0$ und $f : \Omega \to \overline{\mathbb{R}}$ messbar mit $P(f_n \uparrow f) = 1 \Rightarrow$*

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Beweis. Folgt direkt aus der monotonen Konvergenz (Teil (vi) in 10) mit

$$g_n = (f - f_n)\mathbb{1}_{\{\omega \in \Omega : f_n(\omega) \uparrow f(\omega)\}}. \qquad \square$$

Theorem A.30 (Fatou). *Sei $(\Omega, \mathscr{F}, \mu)$ Maßraum $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$* nichtnegativ *und messbar. Dann gilt*

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \liminf_{n\to\infty} f_n d\mu.$$

Das Lemma von Fatou kommt mit den geringsten Voraussetzungen aus (lediglich nicht-negativität, also eine untere Schranke wird gefordert). Allerdings erhält man auch nur eine schwächere Aussage.

Beweis. $k \geq n \Rightarrow f_k \geq \inf_{m \geq n} f_m$, so dass

$$\inf_{k\geq n} \int f_k d\mu \geq \int \inf_{k \geq n} f_n d\mu.$$

Für $n \to \infty$ erhalten wir

$$\liminf_{n\to\infty} \int f_n d\mu \geq \sup_{n\in\mathbb{N}} \int \inf_{k\geq n} f_k d\mu = \int \liminf_{n\to\infty} f_n d\mu$$

wegen monotoner Konvergenz $(\inf_{k\geq n} f_k \uparrow \liminf_{n\to\infty} f_n)$. $\square$

Theorem A.31 (Majorisierte Konvergenz). *Sei* $(\Omega, \mathscr{F}, \mu)$ *Maßraum und* $f, g, (f_n)_{n\geq 1} : \Omega \to \overline{\mathbb{R}}$ *messbar. Sei* $|f_n| \leq g$ *f. s. und* $\lim_{n\to\infty} f_n = f$, $g \in \mathcal{L}^1(\mu)$.

Dann gilt

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

Die Beweisidee ist Fatou auf $g + f$, $g - f$ anzuwenden.

Beweis. Ohne Einschränkung $|f_n| \leq g \ \ \forall \omega \in \Omega \Rightarrow$

(1) $\int (g+f)d\mu \leq \liminf\limits_{n\to\infty} \int (g+f_n)d\mu = \int g d\mu + \liminf\limits_{n\to\infty} \int f_n d\mu$

(2) $\int (g-f)d\mu \leq \int g d\mu - \limsup\limits_{n\to\infty} \int f_n d\mu$, also

$$\int f d\mu \overset{(1)}{\leq} \liminf_{n\to\infty} \int f_n d\mu \leq \limsup_{n\to\infty} \int f_n d\mu \overset{(2)}{\leq} \int f d\mu. \qquad \square$$

Wir definieren für $0 < p < \infty$ die Halbnorm

$$\|f\|_p = \Big(\int |f|^p d\mu \Big)^{1/p}$$

auf dem Raum der p-fach integrierbaren Funktionen

$$\mathcal{L}^p(\mu) = \{ f : \Omega \to \overline{\mathbb{R}} \text{ messbar mit } \|f\|_p < \infty \}$$

Dieser Raum ist auch vollständig, aber da es mehrere Funktionen mit $\|f\|_p = 0$ gibt, sind Limiten nicht eindeutig bestimmt. Man geht deswegen zu dem Quotientenraum[57]

$$L^p(\mu) := \mathcal{L}^p(\mu)/N_p$$

mit $N_p = \{ f : \|f\|_p = 0 \}$ über. Dieser ist ein Banachraum für $p \geq 1$.

Man kann dies auch erweitern auf beschränkte Funktionen $L^\infty(\mu)$ durch die Norm

$$\|f\|_\infty = \inf\{ K : \mu(|f| > k) = 0 \}$$

(ebenfalls ein Banachraum). Der Raum $L^0(\mu)$ besteht schlicht aus allen meßbaren Funktionen.

Unter anderem gilt für $f \in \mathcal{L}^p(\mu)$ und $g \in \mathcal{L}^q(\mu)$, dass[58]

(i) $0 < p, q, r \leq \infty$ und $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$: $\quad \|fg\|_r = \|f\|_p \|g\|_q$

(ii) $1 \leq p \leq \infty$: $\quad \|f + g\|_p \leq \|f\|_p + \|g\|_p$

Für weitere Details zu den L^p-Räumen siehe etwa Werner[59], Kapitel 1.1 Beispiel (h).

[57] Das macht in der Regel keine Probleme. Allerdings ist etwa die Abbildung $f \mapsto f(x_0)$ nicht mehr wohldefiniert.

[58] Die Hölder- und Dreiecksungleichung.

[59] D. Werner. *Funktionalanalysis*. Springer, 2000

Definition A.32 (Konvergenz im p-ten Mittel)**.** Eine Folge $(f_n)_{n\geq 1} \subset \mathcal{L}^p(\mu)$ **konvergiert im p-ten Mittel**, falls

$$\|f_n - f\|_p \xrightarrow[n\to\infty]{} 0.$$

Wir schreiben

$$f_n \xrightarrow[n\to\infty]{L^p(\mu)} f.$$

Konvergiert eine Folge im p-ten Mittel, so auch im q-ten Mittel, falls $q < p$. (Hölder)

Wie bereits erwähnt, ist $L^p(\mu)$, $p \geq 1$ **vollständig** (jede Cauchy-Folge konvergiert) und somit ein Banachraum.

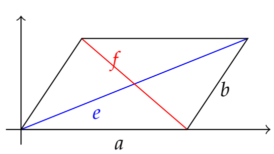
Hilberträume

Auf dem $L^2(\mu)$ hat man eine besondere Struktur, denn dieser Raum ist ein Hilbertraum. Dann lassen sich die allgemeinen Resultate anwenden (siehe: [60]).

[60] D. Werner. *Funktionalanalysis*. Springer, 2000

Lemma A.33 (Parallelogrammidentität)**.** *Ein Banachraum* $(X, \|\cdot\|)$ *ist genau dann ein Hilbertraum falls* $\forall\, f, g$ *gilt, dass*

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2). \qquad \textit{(Parallelogramm)}$$



$2(a^2+b^2)=e^2+f^2$

Beweis. Siehe [61], S.V.1.7. □

[61] D. Werner. *Funktionalanalysis*. Springer, 2000

Satz A.34. *Sei M ein abgeschlossener, linearer Unterraum des Hilbertraums H und $f \in H$. Dann gibt es genau ein $g \in M$, $h \perp M$, so dass*

$$f = g + h.$$

Satz A.35 (Riesz-Fréchet)**.** *Sei H ein Hilbertraum und $F : H \to \mathbb{R}$. F ist genau dann stetig und linear, falls es ein $f \in H$ gibt, so dass*

$$F(h) = \langle f, h\rangle \qquad \forall\, h \in H.$$

Hierbei ist f eindeutig.

Beweis. Siehe [62], V.3.6. □

[62] D. Werner. *Funktionalanalysis*. Springer, 2000

Signierte Maße und der Satz von Radon-Nikodým

Wir wenden uns kurz diesem allgemeineren Begriff zu. Vorstellen kann man sich ein signiertes Maß zum Beispiel als Ladungsverteilung mit positiver und negativer Ladung. Sei $(\Omega, \mathscr{F})$ ein fester Maßraum

Definition A.36. Eine Abbildung $\nu : \mathscr{F} \to \mathbb{R}$ heißt *signiertes Maß*, falls

(i) $\nu(\varnothing) = 0$
(ii) $\nu(\mathscr{F}) = \{\nu(F) : F \in \mathscr{F}\}$ ist entweder Teilmenge von $(-\infty, +\infty]$ oder von $[-\infty, \infty)$.
(iii) Sind $(A_n) \in \mathscr{F}$ p. d., so gilt

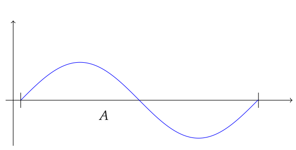
$$\nu\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \nu(A_n).$$

Beispiel A.37. Hat man Maße μ_1, μ_2, wobei eines davon endlich ist, so ist $\nu = \mu_1 - \mu_2$ signiertes Maß. In der Tat hat jedes signierte Maß eine solche Gestalt und μ_1, μ_2 sind bei geeignet „minimaler" Wahl sogar eindeutig.

Definition A.38. Ist ν signiertes Maß und $F \in \mathscr{F}$.

(i) F heißt *(ν-)positiv*, falls $\nu(A) \geq 0$ für alle $\mathscr{F} \ni A \subseteq F$
(ii) F heißt *(ν-)negativ*, falls $\nu(A) \leq 0$ für alle $\mathscr{F} \ni A \subseteq F$
(iii) F heißt *ν-Nullmenge*, falls $\nu(A) = 0$ für alle $\mathscr{F} \ni A \subseteq F$

Lemma A.39. *Ist $\nu : \mathscr{F} \to [-\infty, \infty)$ signiertes Maß und $A \in \mathscr{F}$ mit $\nu(A) \neq -\infty \Rightarrow$ es existiert eine positive Menge P mit $\nu(P) \geq \nu(A)$.*



Beweis. (i) Wir zeigen: Zu jedem $\varepsilon > 0$ gibt es $\mathscr{F} \ni A_\varepsilon \subseteq A$ mit $\nu(A_\varepsilon) \geq \nu(A)$ und $\nu(B) \geq -\varepsilon \; \forall \, \mathscr{F} \ni B \subseteq A_\varepsilon$.

Durch Widerspruch: Angenommen, es existiert $\varepsilon > 0$, so dass die Behauptung falsch ist. Dann enthält jede messbare Menge $C \subseteq A$ mit $\nu(C) \geq \nu(A)$ ein $B \in \mathscr{F}$, so dass $\nu(B) \leq -\varepsilon$. Wir erhalten eine Folge $B_1 \subseteq A$, $B_k \subseteq A \setminus (B_1 \cup \cdots \cup B_{k-1})$ mit $\nu(B_k) \leq -\varepsilon$, $k \geq 1 \Rightarrow \nu(\sum B_i) = -\infty$. Das ist ein Widerspruch zu $\nu(A) > -\infty$.

Wir erhalten demnach eine fallende Folge $(A_{1/n}) \subseteq \mathscr{F}$ und $P := \bigcap A_{1/n}$ ist positiv sowie $\nu(A_{1/n}) \subseteq \nu(A)$, also auch $\lim \nu(A_{1/n}) = \nu(P)$. □

Satz A.40 (Hahn-Zerlegung)**.** *Zu jedem signierten Maß ν existiert eine disjunkte Zerlegung $\Omega = P + N$, in eine positive Menge $P \in \mathscr{F}$ und eine negative Menge $N \in \mathscr{F}$.*

Beweis. Ohne Beschränkung der Allgemeinheit sei $\nu(\mathscr{F}) \subseteq [-\infty, \infty)$ und $\alpha = \sup\{\nu(A) : A \in \mathscr{F}\}$. Nach Lemma A.6 gibt es eine Folge positiver Mengen (A_n) mit $\nu(A_n) \to \alpha$. Dann ist $P = \bigcup_{n \geq 1} A_n$ positiv und $\nu(P) \geq \nu(A_n)$, also $\nu(P) = \alpha$ und somit $\alpha \in \mathbb{R}$.

Nun ist $N = P^c$ negativ: Gäbe es $B \subset N$ mit $N(B) > 0$, so wäre $\nu(P \cup B) > \alpha$. □

Man sieht leicht, dass diese Zerlegung **eindeutig** bis auf Nullmengen ist.

Für ein signiertes Maß ν mit Hahn-Zerlegung $\Omega = P + N$ heißen

$$\left.\begin{aligned} \nu^+(A) &:= \nu(A \cap P) \\ \nu^-(A) &:= \nu(A \cap N) \end{aligned}\right\}, \qquad A \in \mathscr{F}$$

positive (negative) Variation von ν. Da P und N bis auf Nullmengen eindeutig sind, ist $\nu^{+/-}$ wohldefiniert.

Definition A.41. Zwei signierte Maße μ, ν heißen *singulär* (und wir schreiben $\mu \perp \nu$), falls $\Omega = A + A^c$, $A \in \mathscr{F}$ mit $\mu(A) = 0$, $\nu(A^c) = 0$.

Satz A.42 (Jordan-Zerlegung)**.** *Jedes signierte Maß ν erfüllt $\nu = \nu^+ + \nu^-$ mit $\nu^+ \perp \nu^-$. Diese Zerlegung ist* minimal*, d.h.: ist $\nu = \varrho - \sigma$ mit Maßen ϱ, σ, von denen mindestens eines endlich ist, so ist $\nu^+ \leq \varrho$, $\nu^- \leq \sigma$.*

Beweis. Es bleibt lediglich die Minimalität:

$$\nu^+(A) = \nu(A \cap P) = \varrho(A \cap P) - \sigma(A \cap P) \leq \varrho(A \cap P) \leq \varrho(A). \quad \square$$

Beispiel A.43. Sei $f : \Omega \to \overline{\mathbb{R}}$ quasi-integrierbar. Dann ist durch

$$d\mu = f dx$$

ein signiertes Maß definiert und $P = f^{-1}([0, \infty])$, $N = f^{-1}([-\infty, 0))$.

Definition A.44. Wieder betrachten wir den festen Maßraum $(\Omega, \mathscr{F})$

(i) Das Maß ν heißt *absolut stetig* bzgl. μ, falls

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \qquad \text{für alle } A \in \mathscr{F}.$$

Dann schreiben wir $\nu \ll \mu$. Gilt zusätzlich $\mu \ll \nu$, so heißen μ und ν *äquivalent* und wir schreiben $\mu \sim \nu$.

(ii) μ heißt singulär zu ν, falls es $A \in \mathscr{F}$ gibt, so dass

$$\mu(A) = 0 \quad \text{und} \quad \nu(A^c) = 0.$$

Wir schreiben $\mu \perp \nu$.

(iii) Gilt

$$\nu(A) = \int_A f d\mu,$$

für alle $A \in \mathscr{F}$, so heißt f *Dichte* von ν bzgl. μ und wir schreiben

$$\frac{d\nu}{d\mu} = f \quad \text{oder} \quad d\nu = f d\mu.$$

Der Schlüssel zu dem Satz von Radon-Nikodým ist folgendes Lemma:

Lemma A.45. *Sind ν, ϱ endliche Maße mit $\nu \leq \varrho$, so gibt es seine messbare Funktion $h : \Omega \to [0, 1]$ mit*

$$d\nu = h\, d\varrho.$$

Beweis. Für $f \in \mathcal{L}^2(\varrho)$ ist $f \mapsto \int_\Omega f d\nu$ wohldefiniert und stetig ($\mathcal{L}^2$ vollständig). Nach Satz A.35 von Riesz-Fréchet gibt es $h \in \mathcal{L}^2(\varrho)$, so dass

$$\int f d\nu = \langle f, h \rangle_\varrho = \int f h\, d\varrho.$$

Insbesondere ist dies der Fall für $f = 1_A$, $A \in \mathscr{F}$. Man sieht direkt, dass $h(\Omega) \in [0, 1]$ ϱ-f.s. $\square$

Lemma A.46. *Sei μ ein Maß auf $(\Omega, \mathscr{F})$.*

(i) Ist ν σ-finit und $d\nu = f d\mu = g d\mu$, so folgt

$$f = g \qquad \mu\text{-f.s.}$$

(ii) Seien $f : \Omega \to \mathbb{R}_{\geq 0}$ und $g : \Omega \to \mathbb{R}$ messbar. Dann gilt

$$\int g \cdot (f d\mu) = \int (f \cdot g) d\mu.$$

Beweis. (i) Da ν σ-finit ist, gibt es $\Omega_1 \subseteq \Omega_2 \subseteq \cdots \subseteq \mathscr{F}$, so dass $\bigcup_{n\geq 1} \Omega_n = \Omega$ und $\nu(\Omega_n) < \infty$.

Wir setzen $A_n = \Omega_n \cap \{f < g\} \Rightarrow \int_{A_n} (f - g) d\mu = v(A_n) - \nu(A_n) = 0$, also $\mu(A_n) = 0$.

Damit ist

$$\mu(f > g) = \mu\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to\infty} \mu(A_n) = 0.$$

(ii) Offensichtlich ist für $g = 1_A$

$$\int g(f \cdot d\mu) = \int_A f d\mu = \int (1_A \cdot f) d\mu.$$

Die übliche Approximation von messbaren g liefert das Resultat.

□

Beispiel A.47. Natürlich kennen wir bereits Dichten bzgl. des Lebesgue-Maßes (Normalverteilungen etc.). Jede diskrete Verteilung hat allerdings auch eine Dichte bezüglich des Zählmaßes

$$\mu = \sum_{n\geq 1} \delta_{x_n}$$

mit dem Dirac-Maß $\delta_{x_n}(A) := \mathbb{1}_A(x_n)$.

Satz A.48 (Radon-Nikodým)**.** *Es sei μ σ-finites Maß und ν ein signiertes Maß, so dass $\nu \ll \mu$. Dann hat ν eine (μ-f.s.) eindeutige Dichte $f : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ bzgl. μ.*

Beweis. Die Eindeutigkeit folgt aus (f, g quasi-i.b., μ σ-finit)

$$\int_A f d\mu \leq \int_A g d\mu \quad \forall A \in \mathscr{F} \Rightarrow f \leq g \ \mu\text{-f.s.}(\text{betrachte } \underbrace{\{f > g + \tfrac{1}{n}\}}_{\in \mathscr{F}})$$

Nach der Jordan-Zerlegung ist $\nu \ll \mu \Leftrightarrow \nu^+ \ll \mu$ und $\nu^- \ll \mu$, so dass wir ohne Beschränkung der Allgemeinheit annehmen können, dass ν ein Maß ist.

(1) μ, ν endlich $\Rightarrow$
Wir betrachten $\tau = \mu + \nu$. Nach Lemma A.45 existieren h, g so dass $d\mu = g d\tau$ und $d\nu = h d\tau$. Wir setzen $N = \{g = 0\}$, so dass $\mu(N) = 0$, also auch $\nu(N) = 0$, da $\nu \ll \mu$.

Definiere

$$f(x) := \frac{h(x)}{g(x)} \mathbb{1}_{\{x \in N\}}$$

$$\Rightarrow \qquad \nu(A) = \nu(A \cap N^c) = \int_{A \cap N^c} h d\tau$$
$$= \int_{A \cap N^c} f d\mu = \int_A f d\mu$$

(2) μ endlich $\Rightarrow$ Setze

$$\alpha = \sup\{\mu(B) : B \in \mathscr{F}, \nu(B) < \infty\} \qquad (< \infty).$$

Dann gibt es $B_1 \subseteq B_2 \subseteq \cdots \subseteq \mathscr{F}$ mit $\nu(B_n) < \infty$, $\mu(B_n) \to \alpha$, also $\mu(\bigcup_{n \geq 1} B_n) = \alpha$.

Für $A \in \mathscr{F}$, $A \subseteq (\bigcup B_n)^c$, $\nu(A) < \infty$ gilt

$$\alpha + \mu(A) = \lim_{n \to \infty} \mu(B_n \cup A) \leq \alpha \qquad (\nu(B_n \cup A) < \infty),$$

also $\mu(A) = 0$ und somit $\nu(A) = 0$.

Wir erhalten $A \subseteq (\bigcup B_n)^c \Rightarrow \mu(A) = \nu(A) = 0$ oder $\mu(A) > 0$, $\nu(A) = \infty$.

Es ist $E_n := B_n \setminus B_{n-1}$, $E_1 = B_1$, $E := \bigcup E_n$ eine Folge p. d. Mengen mit $\nu(E_n) < \infty \overset{(1)}{\Rightarrow} \exists f_n$ mit $d(\mathbb{1}_{E_n} \nu) = f_n d\mu$

$$\nu = \Big(\sum_{n \geq 1} \mathbb{1}_{E_n} + \mathbb{1}_{E^c} \Big) \cdot \nu$$
$$= \sum_{n \geq 1} f_n d\mu + \infty \cdot \mathbb{1}_{E_c} d\mu = \Big(\sum_{n \geq 1} f_n + \infty \cdot \mathbb{1}_{E^c} \Big) d\mu$$

(3) μ σ-finit
$\Rightarrow$ Es existieren p. d. (Ω_n), $\mu(\Omega_n) < \infty$, so dass $\Omega = \sum_{n \geq 1} \Omega_n$.
Wir definieren $d\mu_n = \mathbb{1}_{\Omega_n} d\mu$, $\nu_n = \mathbb{1}_{\Omega_n} d\nu$
und wenden (2) an $\Rightarrow$ $d\nu_n = f_n d\mu_n$ und $\sum f_n \mathbb{1}_{A_n}$ ist Dichte von ν bzgl. μ. □

Beispiel A.49. Auf σ-Finitheit kann nicht verzichtet werden: $\Omega \neq \emptyset$, $\mathscr{F} = \{\emptyset, \Omega\}$, $\mu(\emptyset) = 0$, $\mu(\Omega) = \infty$, $\nu(\emptyset) = 0$, $\nu(\Omega) = \mathbb{1}$, so ist $\nu \ll \mu$, aber es gibt keine Dichte ($c \cdot \infty \neq \mathbb{1}$).

Als eine erste Anwendung des Satzes von Radon und Nikodým erhalten wir die berühmte Lesbesgue-Zerlegung. Sie erlaubt insbesondere die Zerlegung eines Maßes auf $\mathscr{B}(\mathbb{R})$ in einen absolut stetigen, einen diskreten und einen singulär-stetigen Teil.

Das Maß auf der Cantor Menge ist ein Beispiel eines singulär-stetigen Maßes.

Satz A.50 (Lebesgue-Zerlegung). *Ist μ σ-finites Maß und ν σ-finites, signiertes Maß, so gibt es genau eine Zerlegung*

$$\nu = \varrho + \sigma$$

mit signierten Maßen ϱ, ν, so dass $\varrho \ll \mu$, $\sigma \perp \mu$.

Beweis. Wieder genügt es nach Satz **??**, Maße zu betrachten. Wir setzen $\tau = \mu + \nu$. Nach dem Satz von Radon-Nikodým, Satz A.48, existiert eine Dichte g so dass $d\mu = g d\tau$.

Definiere $N = \{g = 1\}$ und

$$\varrho(A) = \nu(A \cap N^c), \quad \sigma(A) = \nu(A \cap N), \qquad A \in \mathscr{F}. \qquad \square$$

Eindeutigkeit: ÜA

Beispiel A.51. Als Beispiel betrachten wir die Poissonverteilung. Sie hat eine Dichte bezgl. des Zählmaßes

$$d\nu = \sum_{k \geq 0} \delta_{\{k\}}$$

gegeben durch $\nu(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Produkträume

Für uns interessant werden polnische Räume sein, da auf ihnen die zentralen Grenzwertsätze der Stochastik gelten werden. Für einen topologischen Raum $(\Omega, \mathcal{O})$ und für ein $A \subseteq \Omega$ setzen wir

$$A^o := \bigcup \{O \subseteq A : O \in \mathcal{O}\} \qquad \text{(Innere von } A\text{)}$$
$$\overline{A} := \bigcap \{F \supseteq A : F^c \in \mathcal{O}\} \qquad \text{(Abschluss von } A\text{)}$$

Definition A.52. Ein topologischer Raum $(\Omega, \mathcal{O})$ heißt **polnisch**, falls

(i) es eine abzählbare Menge $\Omega' \subseteq \Omega$ gibt, so dass $\overline{\Omega}' = \Omega$,
(ii) er vollständig metrisierbar ist.

Ein Raum heißt *vollständig metrisierbar*, falls es eine Metrik gibt, die die Topologie erzeugt und bzgl. der Ω vollständig (jede Cauchy-Folge konvergiert) ist.

Beispiel A.53. (i) $(\mathscr{C}[0,1], \|\cdot\|_\infty)$ ist polnisch
(ii) jeder **kompakte** metrische Raum ist polnisch
(iii) $(\mathscr{C}[0,\infty), d)$ mit

$$d(f,g) = \sum_{k \geq 1} \frac{1}{2^k} \frac{d_k(f,g)}{1 + d_k(f,g)} \qquad \text{mit } d_k = \sup_{[0,k]} |f-g|$$

ist polnisch (Topologie der gleichmäßigen Konvergenz auf Kompakta).

Für den Satz von Fubini werden wir mehrfach-Integrale betrachten. Diese integrieren über das Produkt von Räumen, was wir nun einführen. Für eine Familie $(\Omega_i)_{i \in I}$ von Mengen heißt

$$\times_{i \in I} \Omega_i := \{(\omega_i)_{i \in I} : \omega_i \in \Omega_i\}$$

Produktraum der $(\Omega_i)_{i \in I}$. Für $H \subseteq J \subseteq I$ definieren wir die **Projektionen** $\pi^J_H : \times_{j \in J} \Omega_j \to \times_{h \in H} \Omega_h$ durch

$$\pi^J_H((\omega_j)_{j \in J}) = (\omega_h)_{h \in H}.$$

Außerdem schreiben wir $\pi^I_H = \pi_H$ und $\pi_i = \pi_{\{i\}}$, $i \in I$.

Definition A.54. Ist $(\Omega_i, \mathcal{O}_i)_{i \in I}$ eine Familie von topologischen Räumen, so heißt die von

$$\left\{ \times_{j \in J} A_j \times \times_{i \in I \setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_j \in \mathcal{O}_j,\ j \in J \right\}$$

erzeugte Topologie *Produkttopologie* auf $\Omega = \times_{i \in I} \Omega_i$.

Bezüglich der Produkttopologie $\mathcal{O}$ sind alle Projektionen π_i, $i \in I$ stetig:

$$\pi_i^{-1}(A_i) = A_i \times \times_{j \in I \setminus \{i\}} \Omega_j \in \mathcal{O}, \qquad A_i \in \mathcal{O}_i.$$

Besonders schöne Eigenschaften erhält man unter Abzählbarkeit.

Satz A.55. *Sei $(\Omega_i, \mathcal{O}_i)_{i \in \mathbb{N}}$ eine Familie polnischer Räume. Dann ist der Produktraum Ω versehen mit der Produkttopologie $\mathcal{O}$ wieder polnisch.*

Beweis. Da Ω_i separabel ist, gibt es abzählbare $\Omega_i' \subseteq \Omega$, so dass $\overline{\Omega}_i' = \Omega$ für alle $i \in I$. Weiterhin bezeichne τ_i die vollständige Metrik, die $\mathcal{O}_i$ erzeugt (über $\{B_\varepsilon(\omega) = \{\omega' \in \Omega : d(\omega, \omega') < \varepsilon\} : \varepsilon > 0, \omega \in \Omega\}$).

Für $\omega, \omega' \in \Omega = \times_{i \in \mathbb{N}} \Omega_i$ wird durch

$$\tau(\omega, \omega') = \sum_{i \geq 1} \frac{1}{\alpha^i} (\tau_i(\omega_i, \omega_i') \wedge 1)$$

eine vollständige Metrik auf Ω definiert, die die Produkt-Topologie $\mathcal{O}$ erzeugt.

Für $\omega' \in \times_{i \in \mathbb{N}} \Omega_i'$ ist

$$B_{\omega'} := \Big\{ \omega \in \times_{i \in \mathbb{N}} \Omega_i' : \omega_i \neq \omega_i' \text{ für endlich viele } i \in \mathbb{N} \Big\}$$

abzählbar und dicht in Ω und die Behauptung folgt. □

Definition A.56. Für eine Familie $(\Omega_i, \mathscr{F}_i)_{i \in I}$ von Maßräumen heißt die σ-Algebra

$$\otimes_{i \in I} \mathscr{F}_i := \sigma\Big(\Big\{ \times_{j \in J} A_j \times \times_{i \in I \setminus J} \Omega_i : J \subseteq I, I \text{ endlich}, A_i \in \mathscr{F}_i \Big\} \Big)$$

die Produkt-σ-Algebra auf $\Omega = \times_{i \in J} \Omega_i$. Ist $\mathscr{F}_i = \mathscr{F}_1 =: \mathscr{F}$, so schreiben wir auch $\mathscr{F}^I = \otimes_{i \in I} \mathscr{F}$.

Wieder sind die Projektionen verträglich mit der Konstruktion: Da

$$\pi_i^{-1}(A_i) = A_i \times \times_{j \in I \setminus \{i\}} \Omega_j \in \otimes_{i \in I} \mathscr{F}_i,$$

ist π_i messbar bzgl. $\otimes_{i \in I} \mathscr{F}_i$. Außerdem ist

$$\otimes_{i \in I} \mathscr{F}_i = \sigma\Big(\Big\{ A_j \times \times_{j \in J \setminus \{i\}} \Omega_j : A_i \in \mathscr{F}_i, i \in I \Big\} \Big).$$

Unter geeigneten Voraussetzungen ist die Borel-σ-Algebra der Produkttopologie gleich der Produkt-σ-Algebren der einzelnen Borel-σ-Algebren. Abzählbarkeit ist eine solche (vgl. Elstrodt, Kapitel III.5).

Lemma A.57. *Ist I abzählbar und $(\Omega, \mathcal{O})$ polnisch, so gilt*

$$\mathscr{B}(\mathcal{O}) = \otimes_{i \in I} \mathscr{B}(\mathcal{O}_i).$$

Beweis: Wir nutzen folgendes, einfaches topologisches Resultat: Ist $(\Omega, \mathcal{O})$ separabel, so hat $\mathcal{O}$ eine abzählbare Basis $\mathcal{B}$, d.h.

$$\mathcal{O} = \left\{ \bigcup_{i \in I} B_i : B_i \in \mathcal{B},\ I \subset \mathbb{N} \right\}$$

Da alle $(\Omega_i, \mathcal{O}_i), i \in I$, separabel sind, haben sie jeweils abzählbare Basen $\mathcal{B}_i$. Dann ist

$$\mathcal{B}' := \left\{ \times_{j \in J} A_j \times \times_{i \in I \setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_i \in \mathcal{B}_i \right\}$$

eine abzählbare Basis von $(\Omega, \mathcal{O})$ und somit $\sigma(\mathcal{B}') = \mathscr{B}(\Omega)$.

Außerdem ist $\mathcal{B}' \subseteq \otimes_{i \in J} \mathscr{B}(\Omega_i)$, also $\mathscr{B}(\Omega) \subseteq \otimes_{i \in I} \mathscr{B}(\Omega_i)$.

Umgekehrt ist für $A_i \in \mathcal{O}_i$

$$A_i \times \times_{j \in I \setminus \{i\}} \Omega_j \in \sigma\left(\left\{ A_i \times \times_{j \in I \setminus \{i\}} \Omega_i : A_i \in \mathcal{O}_i \right\} \right) \subseteq \mathscr{B}(\Omega). \qquad \square$$

Man erhält leicht folgendes Resultat über erzeugende Halbringe:

Lemma A.58. *(i) Sei I endlich, $\mathscr{H}_i$ Halbringe mit $\sigma(\mathscr{H}_i) = \mathscr{F}_i$. Dann ist*

$$\mathscr{H} := \left\{ \times_{i \in I} A_i : A_i \in \mathscr{H}_i, i \in I \right\}$$

ein Halbring mit $\sigma(\mathscr{H}) = \otimes_{i \in I} \mathscr{F}_i$.

(ii) Sei I beliebig, $\mathscr{H}_i$ durchschnittstabil und $\sigma(\mathscr{H}_i) = \mathscr{F}_i$. Dann ist

$$\mathscr{H} := \left\{ \times_{j \in J} A_j \times \times_{i \in I \setminus J} \Omega_i : J \subseteq I, J \text{ endlich}, A_j \in \mathscr{H}_j, j \in J \right\}$$

durchschnittstabil mit $\sigma(\mathscr{H}) = \otimes_{i \in I} \mathscr{F}_i$.

Der Satz von Fubini

Ein wichtiges Hilfsmittel für Maße auf Produkträumen sind **Übergangskerne**. Sie beschreiben in einem dynamischen Kontext Übergangswahrscheinlichkeiten von so genannten Markov-Prozessen, wie wir gleich in einem Beispiel sehen werden.

Definition A.59. Seien $(\Omega_1, \mathscr{F}_1)$ und $(\Omega_2, \mathscr{F}_2)$ Messräume. Eine Abbildung $\kappa : \Omega_1 \times \mathscr{F}_2 \to \mathbb{R}_{\geq 0}$ heißt *Übergangskern*, falls

(i) für alle $\omega_1 \in \Omega_1$ ist $\kappa(\omega_1, \cdot)$ ein Maß auf $(\Omega_2, \mathscr{F}_2)$
(ii) für alle $A_2 \in \mathscr{F}_2$ ist $\kappa(\cdot, A_2)$ $\mathscr{F}_1$-messbar.

Ein Übergangskern heißt σ-finit, falls es $A_1, A_2, \ldots \uparrow \Omega_2$ gibt mit $\sup_{\Omega_1} \kappa(\omega_1, A_n) < \infty$ für alle $n = 1, 2, \ldots$

Er heißt *stochastischer Kern* oder *Markovscher Kern*, falls

$$\kappa(\omega_1, \Omega_2) = 1 \qquad \text{für alle } \omega_1 \in \Omega_1.$$

Beispiel A.60 (Markov-Kette)**.** Sei $\Omega = \{\omega_1, \ldots, \omega_n\}$ mit Wahrscheinlichkeiten $(p_{ij})_{1 \leq i,j \leq n}$, so dass mit $\sum_{j=1}^n p_{ij} = 1 \forall i$. Dann ist

$$\kappa(\omega_i, \cdot) := \sum_{j=1}^n p_{ij} \delta_{\omega_j}$$

ein stochastischer Kern von $(\Omega, \mathscr{P}(\Omega))$ nach $(\Omega, \mathscr{P}(\Omega))$. Die (p_{ij}) bilden die Übergangswahrscheinlichkeiten einer homogenen **Markov-Kette**.

Ein typisches Beispiel kann man sich so vorstellen. Betrachten wir als Zustandsraum $\Omega = \mathbf{Z}$. Ausgehend von einem Punkt x geht die Markovkette entweder einen Schritt nach oben, oder einen nach unten, so dass der Übergangskern

$$\kappa(x, \cdot) = p\delta_{x+1} + (1-p)\delta_{x-1}$$

ist. Die Frage ist nun: Reicht die Angabe dieser Übergangswahrscheinlichkeiten um die Verteilung des Markov Prozesses genau zu beschreiben.

Lemma A.61 (Messbarkeit integrierbarer Schritte)**.** *Sei κ ein σ-finiter Übergangskern und $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_{\geq 0}$ $\mathscr{F}_1 \otimes \mathscr{F}_2$-messbar. Dann ist*

$$\omega_1 \mapsto \int f(\omega_1, \omega_2) \kappa(\omega_1, d\omega_2)$$

$\mathscr{F}_1$-messbar.

Beweis (Skizze). Zunächst nehmen wir $\kappa(\omega_1, \Omega_2) < \infty$ an. (Dann Erweiterung mittels $A_1, A_2 \uparrow \Omega_2$.) Wir betrachten

$$\mathscr{D} = \left\{ A \in \mathscr{F}_1 \otimes \mathscr{F}_2 : \omega_1 \mapsto \int \mathbb{1}_A(\omega_2)\kappa(\omega_1, d\omega_2) \text{ ist } \mathscr{F}_1\text{-messbar} \right\}$$

Dies ist ein durchschnittstabiles Dynkin-System mit $\mathscr{H} = \{A_1 \times A_2, A_i \in \mathscr{F}_i\} \subseteq \mathscr{D}$. Nun gilt $\sigma(\mathscr{H}) = \mathscr{F}_1 \otimes \mathscr{F}_2$, also $\sigma(\mathscr{H}) = \mathscr{F}_1 \times \mathscr{F}_2 \subseteq \mathscr{D} \subseteq \mathscr{F}_1 \otimes \mathscr{F}_2$ und die Aussage folgt für $f = \mathbb{1}_A$, $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$.

Wir erhalten die Aussage für Treppenfunktionen und mit monotoner Konvergenz für alle messbaren $f \geq 0$. □

Theorem A.62 (Ionescu-Tulcea). *Seien $(\Omega_i, \mathscr{F}_i)$, $i = 0, \ldots, n$ Messräume, μ ein σ-finites Maß auf $(\Omega_0, \mathscr{F}_0)$ und κ_i σ-finite Übergangskerne von $(\times_{j=0}^{i-1}\Omega_j, \otimes_{j=0}^{i-1}\mathscr{F}_j)$ nach $(\Omega_i, \mathscr{F}_i)$, $i = 1, \ldots, n$. Dann gibt es genau ein σ-finites Maß $\kappa = \mu \otimes_{i=1}^{n} \kappa_i$ auf $(\times_{j=0}^{n}\Omega_j, \otimes_{j=0}^{n}\mathscr{F}_j)$, so dass*

$$\kappa(A_0 \times \cdots \times A_n) \int_{A_0} \cdots \int_{A_n} \kappa_n(\omega_0, \ldots, \omega_{n-1}, d\omega_n) \cdots \kappa_1(\omega_0, d\omega_1)\mu(d\omega_0). \tag{A.63}$$

Beweis. Wir beweisen die Aussage für $n = 1$, der allgemeine Fall folgt dann per Induktion. Der Beweis ist eine typische Anwendung des Maßfortsetzungssatzes: Wir starten mit dem durchschnittstabilen Halbring

$$\mathscr{H} = \left\{ \times_{i=0}^{n} A_i : A_i \in \mathscr{F}_i \right\}$$

und betrachten die durch (A.63) definierte Mengenfunktion κ auf $\mathscr{H}$.

(i) *κ ist σ-finit:* Seien $(\Omega_n^i) \subseteq \mathscr{F}_i$ so dass $(\Omega_n^i) \uparrow \Omega^i$, $i = 0, 1$ und $\mu(\Omega_n^i) < \infty$ und $\sup_{\omega^0 \in \Omega^0} \kappa_1(\omega_0, \Omega_n^i) =: C_n < \infty$.

$$\Rightarrow \qquad \begin{aligned} \kappa(\Omega_n^0 \times \Omega_n^1) &= \int_{\Omega_n^0} \int_{\Omega_n^0} \kappa(\omega^0, d\omega^1)\mu(d\omega^0) \\ &\leq C_n \mu(\Omega_n^1) < \infty. \end{aligned}$$

Da $\Omega_n^0 \times \Omega_n^1 \uparrow \Omega_0 \times \Omega_1$ ist κ σ-finit.

(ii) *κ ist σ-additiv:* Für $A_1, A_2 \in \mathscr{H}$ p. d. mit $A = \sum_{i=1}^{\infty} A_i \in \mathscr{H}$ ist

$$\kappa(A) = \int \underbrace{\int \mathbb{1}_A(\omega_0,\omega_1)\kappa_1(\omega_0, d\omega_1)}_{\kappa_1(\omega_0,\cdot) \text{ ist Maß (also } \sigma\text{-subadditiv)}} \mu(d\omega_0)$$

$$= \int \sum_{n\geq 1} \underbrace{\int \mathbb{1}_{A_n}(\omega_0,\omega_1)\kappa_1(\omega_0, d\omega_1)}_{\geq 0} \mu(d\omega_0)$$

$$\overset{(*)}{=} \sum_{n\geq 1} \underbrace{\int\int \mathbb{1}_{A_n}(\omega_0,\omega_1)\kappa_1(\omega_0, d\omega_1)\mu(d\omega_0)}_{\kappa(A_n)}$$

$$(*) := \text{ monot. Konv.}$$

Mit dem Maßfortsetzungssatz erhalten wir die Behauptung. □

Bemerkung A.64. Mit etwas mehr Aufwand kann man die Aussage auch für $n = \infty$ beweisen, siehe [63].

[63] V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1991

Theorem A.65 (Satz von Fubini)**.** *Seien $(\Omega_i, \mathscr{F}_i)$, μ_0, κ_i und $\kappa = \mu_0 \otimes_{i=1}^n \kappa_i$ wie in Theorem 71 gegeben und $f : \times_{i=0}^n \Omega_0 \to \mathbb{R}_{\geq 0}$ messbar bzgl. $\otimes_{i=0}^n \mathscr{F}_i$. Dann gilt*

$$\int f d\kappa = \int \Big(\cdots \Big(\int f(\omega_0, \ldots, \omega_n)\kappa_n(\omega_0, \ldots, \omega_{n-1} d\omega_n) \Big) \cdots \kappa_1(\omega_0, d\omega_1) \Big) \mu(d\omega_0). \tag{A.66}$$

Die Aussage gilt auch für beliebig messbare $\mathscr{F}$ mit $\int |f| d\kappa < \infty$.

Beweis. Für alle $A \in \otimes_{i=0}^n \mathscr{F}_i$ betrachten wir $A \mapsto \int_A d\kappa =: I(A)$.

Für alle $A \in \left\{ \times_{i=0}^n A_i : A_i \in \mathscr{F}_i \right\}$ gilt $I(A) = \kappa(A)$ und somit (A.66).

Wegen der Linearität und der monotonen Konvergenz folgt (A.66) für alle nicht-negativen, messbaren Funktionen.

Messbarkeit der einzelnen Integrale folgt hierbei nach Lemma 70. □

Bemerkung A.67. Die Aussage für Produktmaße $\mu_0 \otimes \cdots \otimes \mu_n$ erhält man als einfachen Spezialfall!

Beispiel A.68 (Mehrdimensionales Lebesgue-Maß)**.** Mit $\lambda_d = \otimes_{i=1}^d \lambda$ bezeichnen wir das Lebesgue-Maß auf dem $\mathbb{R}^d$

$$f(x,y) = \frac{xy}{(x^2+y^2)^2} \Rightarrow \int f(x,y) dy = 0 \quad \forall x$$

$$\Rightarrow \qquad \iint f(x,y)\lambda(dy)\lambda(dx) = \iint f(x,y)dx)\lambda(dy) = 0.$$

Allerdings ist $|f|$ nicht bzgl. λ_2 integrierbar!

Aus Gleichheit und Endlichkeit der Mehrfachintegrale kann man also nicht auf die Integrierbarkeit des Integrand schließen.

Betrachten wir zwei unabhängige Zufallsvariable X und Y, so heißt die Verteilung von $X+Y$ die **Faltung**.

Die Faltung ist demnach die Verteilung $P(X+Y \leq z)$ für alle $z \in \mathbb{R}$.

Wir erinnern noch kurz an die Definition A.25 wo wir das Bildmaß (oder den push-forward) eines Maßes unter einer Abbildung eingeführt haben. Dies kann man nutzen um etwa die Faltung präzise einzuführen.

Definition A.69. Seien $\mu_1, \dots, \mu_n$ σ-finite Maße auf $\mathscr{B}(\mathbb{R})$, $\mu = \otimes_{i=1}^n \mu_i$ deren Produktmaß und $S(x) = x_1 + \cdots + x_n$, $x \in \mathbb{R}^n$. Dann heißt

$$S_* \mu := \mu_1 * \cdots * \mu_n$$

die *Faltung* der Maße $\mu_1, \dots, \mu_n$.

Aufgabe 7. Haben die Maße μ_i Dichten f_i bzgl. λ, $i = 1, 2$ und setzt man

$$f(t) := \int f_1(s) f_2(t-s) \lambda(ds),$$

so gilt

$$\mu_1 * \mu_2 = f_{\mu * \nu} \cdot \lambda. \qquad \text{(Fubini)}$$

Betrachte $\quad P(X+Y \leq x) = P(X \leq x - Y)!$

Aufgabe 8. Die Faltung zweier Normalverteilungen ist wieder normalverteilt.

Der Existenzsatz von Kolmogorov

Der Existenzsatz von Kolmogorov ist der zentrale Grund, warum wir im Folgenden stets polnische Räume betrachten werden.

Bisher haben wir endliche bzw. abzählbare Produkträume betrachtet, was für unsere Anwendungen nicht ausreichen wird. Wenn wir an die Brownsche Bewegung denken - also an eine stetige Funktion, so werden wir auch an allgemeineren Situationen interessiert sein.

Der Satz von Kolmogorov erlaubt die Erweiterungen auf größere Räume, kommt allerdings nicht ohne die Voraussetzung aus, dass diese *polnisch* sind.

Wir erinnern daran, dass ein topologischer Raum $(\Omega, \mathcal{O})$ *polnisch* heißt, falls er vollständig metrisierbar ist und separabel.

- vollständig metrisierbar: Es gibt eine Metrik r so dass $\mathcal{O}$ von r erzeugt wird, d.h. für jedes $A \in \mathcal{O}$ gibt es ein B_ε, so dass $B_\varepsilon \subseteq A$ ist. ($B_\varepsilon(\omega) = \{\omega' \in \Omega : r(\omega, \omega') < \varepsilon\}$) <u>und</u> dass (Ω, r) vollständig ist (jede Cauchy-Folge konvergiert)
- separabel: Es gibt eine abzählbare Teilmenge Ω', so dass

$$\overline{\Omega}' = \bigcap\{A \supseteq \Omega' : A^c \in \mathcal{O}\} = \Omega.$$

Beispiel A.70. (i) $\mathbb{R}^d = \overline{\mathbb{Q}}^d$ ist separabel und vollständig
(ii) Sei $K \subseteq \mathbb{R}$ kompakt $\Rightarrow (C(K), \|\cdot\|_\infty)$ ist **polnisch** (Weierstraß'scher Approximationssatz – Polynome)

Wie bisher bezeichnen wir $\mathscr{F}^J = \otimes_{j \in J}$. Wir verwenden im Folgenden die praktische Konvention

$$J \Subset I := \{ J \subseteq I : \ J \text{ endlich } \}.$$

In diesem Sinne ist etwa $(P_J)_{J \Subset I} = (P_J : J \subseteq I, J \text{ endlich})$ zu verstehen.

Definition A.71. Sei $(\Omega, \mathscr{F})$ ein Maßraum und I eine beliebige Indexmenge. Eine Familie von Wahrscheinlichkeitsmaßen $(P_J)_{J \Subset I}$ heißt *projektive Familie*, falls P_J Wahrscheinlichkeitsmaß auf $(\Omega^J, \mathscr{F}^J)$ ist und

$$P_H = \pi^J_{H*} P_J$$

für alle $H \subseteq J \subseteq I$, J endlich.

Definition A.72. Existiert für eine projektive Familie $(P_J)_{J \Subset I}$ ein Wahrscheinlichkeitsmaß P_I auf $\mathscr{F}^I = \otimes_{i \in I} \mathscr{F}_i$ mit $P_J = \pi_{J*} P_I$ für alle $J \subseteq I$, J endlich, so heißt P_I *projektiver Limes* der Familie P. Wir schreiben

$$P_I = \varprojlim_{J \Subset I} P_J.$$

In mindestens zwei Situationen spielen projektive Familien eine große Rolle.

Beispiel A.73. Sei $(\Omega, \mathscr{F}, P)$ ein Wahrscheinlichkeitsraum und I eine unendliche Indexmenge. In Theorem 71 haben wir für jedes $J \Subset I$ das Produktmaß $P^{\otimes J}$ auf $\mathscr{F}^J$ definiert.

Zentrale Beobachtung ist, dass die Familie $(P^{\otimes J})_{J \Subset I}$ projektiv ist: Ist nämlich $H \subseteq J \Subset I$, so ist für $A_i \in \mathscr{F}, i \in H$,

$$\begin{aligned}
(\pi^J_H)_* P^{\otimes J}\big(\times_{i \in H} A_i\big) &= P^{\otimes J}\big((\pi^J_H)^{-1}\big(\times_{i \in H} A_i\big)\big) \\
&= P^{\otimes J}\big(\times_{i \in H} A_i \times \times_{i \in J \setminus H} \Omega\big) \\
&= \prod_{i \in H} P(A_i) \cdot \prod_{i \in J \setminus H} P(\Omega) \\
&= \prod_{i \in H} P(A_i) \\
&= P^{\otimes H}\big(\times_{i \in H} A_i\big).
\end{aligned}$$

Allerdings haben wir noch nicht gezeigt, wann es den projektiven Limes von $(P^{\otimes J})_{J \Subset I}$ gibt. Diesen würden wir dann das unendliche Produktmaß $P^{\otimes I}$ nennen.

Beispiel A.74. Sei $(\Omega, \mathscr{F}, P)$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, $(\Omega', \mathscr{F}')$ ein Messraum und $X_i : \Omega \to \Omega', i \in I$ eine Familie von Zufallsvariable. Wir werden eine solche Familie $X := (X_i)_{i \in I}$ einen *stochastischen Prozess* nennen.

So ist z.B. $W = (W_t)_{t \geq 0}$ - die Brownsche Bewegung - ein stochastischer Prozess, d.h. eine Familie von Zufallsvariablen.

Dann ist formal gesehen $X : \Omega \to \Omega'^I$ mit $X(\omega) = (X_i(\omega))_{i \in I}$. Man kann sich nun fragen, ob die Verteilung von X (d.h. das Bildmaß $X_* P$) als Verteilung auf $\mathscr{F}'^I$ existiert.

Hierzu sei bemerkt, dass $P'_J := ((X_j)_{j \in J})_* P, J \Subset I$ eine projektive Familie ist. Ist nämlich $H \subset J \Subset I$ und $\widetilde{A}_i \in \mathscr{F}', i \in H$, dann ist

$$\begin{aligned}
(\pi^J_H)_* P'_J \big(\times_{j \in H} \widetilde{A}_j \big) &= P'_J \big((\pi^J_H)^{-1} \times_{j \in H} \widetilde{A}_j \big) \\
&= P'_J \big(\times_{j \in H} \widetilde{A}_j \times \times_{j \in J \setminus H} \Omega' \big) \\
&= P \big(X_j \in \widetilde{A}_j, j \in H \text{ und } X_j \in \Omega', j \in J \setminus H \big) \\
&= P \big(X_j \in \widetilde{A}_j, j \in H \big) \\
&= P'_H \big(\times_{j \in H} \widetilde{A}_j \big).
\end{aligned}$$

Wie im Folgenden gezeigt wird, gibt es die Verteilung $X_* P$ (was dann der projektive Limes der $(P'_J)_{J \Subset I}$ ist) zumindest dann, wenn $\mathscr{F}'$ die Borel'sche σ-Algebra eines polnischen Raumes ist.

Bemerkung A.75 (Eindeutigkeit des projektiven Limes)**.** Seien P_I und P'_I zwei projektive Limiten, so ist für $A := \times_{i \in J} A_i \times \times_{i \in I \setminus J} \Omega_i \in \mathscr{H}$ mit $\mathscr{H}$ aus Lemma A.58 und $J \Subset I$

Zu jeder projektiven Familie gibt es höchstens *einen* projektiven Limes.

$$P_I(A) = P_J\Big(\times_{i \in J} A_i \Big) = P'_J\Big(\times_{i \in J} A_i \Big) = P'_I(A).$$

Damit stimmen P_I und P'_I auf dem schnittstabilen Erzeuger überein und nach Satz A.16 gilt $P_I = P'_I$.

Das folgende, berühmte Theorem, welches unabhängig voneinander von P.J. Daniell (welcher abzählbares I betrachtet hat [64]) und A.N. Kolmogorov () gezeigt wurde besagt nun, dass bei polnischen Räumen ein solcher Limes stets existiert.

[64] Percy John Daniell. Functions of limited variation in an infinite number of dimension. *Annals of Mathematics*, pages 30–38, 1919

Theorem A.76 (Existenz von Prozessen)**.** *Sei* $(\Omega, \mathcal{O})$ *ein* polnischer *Raum,* $\mathscr{F} = \mathscr{B}(\mathcal{O})$ *und* $(P_J)_{J \Subset I}$ *eine projektive Familie von Wahrscheinlichkeitsmaßen auf* $\mathscr{F}$*. Dann existiert der projektive Limes*

$$P_I := \varprojlim_{J \Subset I} P_J.$$

Beweis. Sei $\mathscr{H}$ wie in Lemma A.58 und μ eine additive Mengenfunktion auf dem Halbring $\mathscr{H}$, definiert durch

$$\mu\Big(\times_{j\in J} A_j \times \times_{i\in I\setminus J}\Omega\Big) := P_J\Big(\times_{j\in J} A_j\Big).$$

Damit ist $\mathscr{H}$ ein Halbring ist und μ ein wohldefinierter Inhalt auf $\mathscr{H}$.

Für die σ-Additivität von μ muss man etwas weiter ausholen und sich Kompaktheitsargumente zu nutze machen, die wir hier zunächst ausklammern. (Für einen leicht andern Beweis siehe Klenke [65]).

[65] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2013

Weiterhin ist $\mu(\Omega^I) = 1$, also lässt μ mit dem Maßfortsetzungssatz, Theorem A.22, in eindeutiger Weise auf ein Maß P_I auf $\sigma(\mathscr{H}) = \mathcal{F}^I$ fortsetzen. Dies muss wegen Eindeutigkeit der projektive Limes von $(P_J)_{J\Subset I}$ sein. $\square$

Sonstiges

Das Banach-Tarski Paradoxon

Das Banach-Tarski Paradoxon besagt, dass man eine disjunkte Zerlegung einer Kugel finden kann, deren Teile man zu zwei Kugeln zusammensetzen kann.

Damit ist eine vernünftige Konstruktion eines Maßes auf allen Teilmengen ab der Dimension 3 nicht mehr möglich - man muss sich auf meßbare Mengen konzentrieren. Der Beweis dieser Aussage ist nicht einfach, auf Wikipedia findet sich eine schöne Darstellung sowie in dem Buch [66].

...

[66] Leonard M Wapner, Harald Höfner, and Brigitte Post. *Aus 1 mach 2: wie Mathematiker Kugeln verdoppeln*. Springer, 2008

Theorem A.77 (Banach-Tarski)**.** *If X and Y are bounded subsets of $\mathbb{R}^3$ having nonempty interiors, then there exist a natural number n and partitions $\{X_j \colon 1 \leq j \leq n\}$ and $\{Y_j \colon 1 \leq j \leq n\}$ of X and Y, respectively, such that X_j, is congruent to Y_j for all j.*

Zwei Mengen heißen hierbei kongruent, wenn sie durch eine Isometrie ineinander überführt werden können. Für weitere Details sei auf [67] verwiesen

[67] Karl Stromberg. The banach-tarski paradox. *The American Mathematical Monthly*, 86(3):151–161, 1979

Literaturverzeichnis

[1] Patrick Billingsley. *Probability and Measure*. Wiley, 3 edition, 1995.

[2] Fisher Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[3] V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1991.

[4] C. Czado and T. Schmidt. *Mathematische Statistik*. Springer Verlag. Berlin Heidelberg New York, 2011.

[5] Percy John Daniell. Functions of limited variation in an infinite number of dimension. *Annals of Mathematics*, pages 30–38, 1919.

[6] Jürgen Elstrodt. *Maß-und Integrationstheorie*. Springer, 2018.

[7] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter, Berlin, 2011.

[8] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter, Berlin, 2011.

[9] Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, 1977.

[10] Norbert Henze. *Stochastik für Einsteiger*. Springer, 2013.

[11] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2013.

[12] Ludger Rüschendorf. *Wahrscheinlichkeitstheorie*. Springer, 2016.

[13] Karl Stromberg. The banach-tarski paradox. *The American Mathematical Monthly*, 86(3):151–161, 1979.

[14] Leonard M Wapner, Harald Höfner, and Brigitte Post. *Aus 1 mach 2: wie Mathematiker Kugeln verdoppeln*. Springer, 2008.

[15] D. Werner. *Funktionalanalysis*. Springer, 2000.